THE PHYSICAL SOCIETY

REPORTS ON PROGRESS IN PHYSICS

VOLUME XVIII (1955)

EXECUTIVE EDITOR
A. C. STICKLAND

Price to Non-Fellows £2 10*s. net* (*Postage* 1*s.* 6*d.*)
Price to Fellows (*one copy*) 27*s.* 6*d. net* (*Postage* 1*s.* 6*d.*)

Published by
THE PHYSICAL SOCIETY
1 Lowther Gardens, Prince Consort Road
London S.W.7
1955

Printed by Taylor & Francis Ltd.
Red Lion Court, Fleet Street, London, E.C.4

CONTENTS

	PAGE
Preface	iv
The Displacement of Atoms in Solids by Radiation. By G. H. KINCHIN and R. S. PEASE	1
Ionization by Relativistic Particles. By B. T. PRICE	52
Methods of Measuring Strong Magnetic Fields. By J. L. SYMONDS	83
Theory of Radiation. By J. C. GUNN	127
Magnetostriction and Magnetomechanical Effects. By E. W. LEE	184
Electrostriction. By H. F. KAY	230
Magnetic Cooling. By E. AMBLER and R. P. HUDSON	251
Paramagnetic Resonance II. By K. D. BOWERS and J. OWEN	304
The Ionization and Dissociation of Complex Molecules by Electron Impact. By J. D. CRAGGS and C. A. MCDOWELL	374
A Survey of Field Theory:	
Part I. General Background. By R. E. PEIERLS	424
Part II. Fields and Particles. By A. SALAM	433
Part III. Interaction of Quantized Fields. By P. T. MATTHEWS	441
Part IV. Properties of the π–Meson. By P. T. MATTHEWS	452
Part V. High Energy Experiments. By G. FELDMAN	462
Part VI. Final Remarks. The New Particles. By R. E. PEIERLS	470

PREFACE

The *Reports on Progress in Physics* is under the general supervision of the Papers Committee of the Physical Society. The Members of the Papers Committee were as follows:

Chairman:—Professor S. Devons.

Honorary Papers Secretary:—Dr. H. H. Hopkins.

Ordinary Members:—The President (Professor H. S. W. Massey), Professor W. E. Burcham, Dr. E. H. S. Burhop, Mr. C. H. Collie, Dr. W. Ehrenberg, Dr. F. C. Frank, Mr. D. W. Fry, Professor F. Llewellyn Jones, Professor G. O. Jones, Professor H. Jones, Dr. H. G. Kuhn, Professor R. E. Peierls, Dr. R. W. B. Stephens, Dr. L. C. Thomson, Dr. H. Wilman.

Reports published in the Volume are also being issued as individual articles, at prices depending on their length. Information may be obtained from the offices of the Society.

A. C. STICKLAND
Secretary-Editor

THE DISPLACEMENT OF ATOMS IN SOLIDS BY RADIATION

By G. H. KINCHIN and R. S. PEASE

Atomic Energy Research Establishment, Harwell, Didcot, Berks

CONTENTS

PAGE

Synopsis 2

§ 1. Introduction 4
1.1. General 4
1.2. Historical 4
1.3. Relation between theory and experiment 5

§ 2. The production of displaced atoms 5
2.1. Energy considerations 5
2.1.1. Displacement energy 5
2.1.2. Threshold radiation energies 6
2.2. The number of displaced atoms 6
2.2.1. Moving atoms: principles 7
2.2.2. Moving atoms: hard sphere collisions 9
2.2.3. Moving atoms: Rutherford collisions 10
2.2.4. Moving atoms slowing down to rest 10
2.2.5. Fast neutrons 11
2.2.6. Electrons 12
2.2.7. Gamma rays 13
2.3. Distribution of displaced atoms and vacancies 14
2.3.1. Collision models 14
2.3.2. 'Displacement spikes' and 'thermal spikes' 15
2.4. Effects in compounds 16
2.4.1. Replacement collisions 16
2.4.2. Disordering 17

§ 3. Recovery 17
3.1. Recombination of interstitials with vacancies 17
3.1.1. Close pairs of interstitials and vacancies 17
3.1.2. Random distribution of interstitials and vacancies 18
3.2. Competing processes 19
3.3. Accumulation of damage 19
3.4. Saturation 20
3.5. Radiation annealing 21

§ 4. Experimental irradiation techniques 21
4.1. Reactor irradiation 21
4.1.1. Fast neutrons 21
4.1.2. Thermal neutron reactions 22
4.2. Cyclotron irradiation 23
4.3. Van de Graaff irradiation 23
4.4. Gamma irradiation 23
4.5. Natural radioactivity 23

§ 5. Experimental results 25
5.1. Metals 25
5.1.1. Copper 25
5.1.2. Copper–gold 29
5.1.3. Other metals and alloys 31

PAGE
5.2. Semiconductors 33
5.2.1. Germanium 33
5.2.2. Silicon 36
5.2.3. Indium antimonide 36
5.2.4. Graphite 37
5.3. Covalent compounds 37
5.3.1. Quartz 37
5.3.2. Thermal, optical and magnetic properties 41
5.3.3. X-ray studies 43
5.3.4. The metamict state 46
5.4. Alkali halides 47
Acknowledgments 48
References 48

SYNOPSIS

Two questions are raised by phenomena observed in irradiated solids. What kind of defects are produced? How do such defects alter the properties of the solids? For both questions the answers, in so far as they are known, vary widely from one type of solid to another and, rather less markedly, from one type of radiation to another. In molecular compounds and plastics, the changes observed are largely a consequence of chemical rearrangement induced by ionization. In ionic compounds also, the majority of the effects studied arise from ionization, although the processes involved have been thought of in physical rather than in chemical terms. In metals and covalent compounds, large changes are induced by the displacement of atoms from their lattice sites by collision processes, and the defects produced are primarily interstitial–vacancy pairs, that is, Frenkel defects. This article is mainly concerned with this last case.

Theoretical and experimental work shows that about 25 ev has to be transferred to an atom to form a Frenkel defect. When an atom is displaced by direct collision with the bombarding radiation, we refer to it as a primary knock-on. In some cases, notably electron irradiation, the majority of the displaced atoms are primary knock-ons. In other cases, notably fast neutron irradiation, the majority of the displaced atoms are produced as a consequence of the collisions made by the primary knock-ons in slowing down to rest. With charged particle irradiation, the Coulomb force between the nucleus of the stationary atom and the charged particle is largely responsible for producing the primary knock-ons. For sufficiently high energy radiation, the collisions are ordinary unscreened Rutherford collisions; in an intermediate range of energies, the simple Coulomb interaction has to be modified to allow for the screening effect of the extra-nuclear electrons; and at the lowest energies the Coulomb forces and the screening electrons produce, to a first approximation, hard sphere interaction. In many cases, the collisions made by the primary knock-ons come into this last category. Fast neutron bombardment produces primary knock-ons by nuclear collisions. The displaced atoms which have been detected in γ-irradiated metals are thought to be produced by Compton electrons.

It is insufficient in most cases to suppose that the defects formed are randomly distributed interstitial atoms and vacancies. The majority of the displaced atoms cannot move more than a few interatomic distances from their original lattice sites before coming to rest. Also, the defects are often concentrated into small volumes; in some cases this grouping is so marked that the production of the defects is probably best thought of as a consequence of rapid melting and solidification of the small volume.

For most room temperature irradiations, some recovery takes place during and immediately after irradiation. This must usually be the recombination of close pairs of interstitials and vacancies, and it occurs however low the dose. Consequently the temperature of irradiation not only controls the ultimate concentration of defects obtainable at high doses, but also affects the rate of accumulation of defects at low doses. Also, the recovery reactions often have a high effective order of reaction, so that although some recovery may occur within a few hours, all the effects may be removed only when the temperature of the specimen is raised.

Experimental studies of the effects in metals have mainly used low temperature irradiations, since at room temperature recovery is so marked that only small effects, and sometimes none, can be detected; copper, for instance, shows marked recovery at 40°K. The process of displacement and recovery can, however, produce disorder in an ordered alloy at room temperature, as has been observed in several instances. The electrical resistivity of metals is increased by irradiation because the electron mean free paths are reduced by the defects, and it provides an easily measured parameter that has been much used for studying irradiation and recovery processes. Activation energies of the recovery processes have been measured and compared with activation energies obtained from calculations or observed in other processes. The simple recombination of interstitials and vacancies is not adequate to explain all the stages of recovery observed in the resistivity of metals, and more complicated defects must be formed. There is as yet no real evidence for the nature of these defects. This is particularly true of the stage associated with the recovery of the increase of yield stress caused by irradiation.

The electrical properties of semiconductors, which change dramatically at quite low doses, have been extensively studied. In this case, the changes arise from the trapping of charge carriers at the defects. In graphite, which has properties intermediate between a metal and a semiconductor, the changes of electrical resistance arise both from mean free path changes and also from trapping of charge carriers.

In insulators, thermal conductivity at low temperatures is very sensitive to defects; moreover, by measuring the thermal conductivity at different temperatures it has been possible to distinguish between isolated defects of atomic dimensions, and large groups of defects which give phonon scattering analogous to boundary scattering. X-ray measurements have shown that lattice expansion is a characteristic irradiation effect, and there is a rough correlation between the expansion and the proportion of interstitial atoms. Measurable damage to the lattice periodicity sometimes accompanies the expansion. In most covalent compounds the defects appear to be more stable than in metals, for measurable effects are easily obtained at room temperature, and sometimes heating at 1000°C fails to produce complete recovery. In many cases it seems that the end product of prolonged bombardment is a saturation concentration of Frenkel defects of not more than about 1% in the original crystal lattice. In certain cases, however, the crystal structure is broken down, and the end product is a glassy amorphous state known for many years to mineralogists as the metamict state. Many radioactive minerals, which have had very large irradiation doses, are found in this state, though they retain a relatively perfect outward morphology, and can be restored to a crystalline state by heating at high temperatures. It is not understood why the metamict state appears in some materials and not in others.

§ 1. Introduction

1.1. *General*

When solids are bombarded with radiation, electrons may be removed from their normal orbits, and atoms may be knocked out of their normal sites. Further, impurities may be introduced, either by nuclear transmutations or by bombarding ions coming to rest in the material. As a result, the irradiation of solids is apt to change their properties, sometimes slightly, sometimes drastically. Much of the current interest in these *irradiation effects* has been aroused by the need to understand and mitigate them in materials used in nuclear reactors (Wigner 1946); in this context, the effects are often referred to as *irradiation damage*. However, with the increased understanding of the role of defects in crystalline solids, considerable interest has also been aroused by the possibility of the controlled introduction of defects by irradiation, which can have valuable applications.

This article is primarily concerned with the displacement of atoms, and the consequent effects. The displacement of electrons produces effects in insulators, most notably coloration effects in alkali halides, a subject recently reviewed by Seitz (1954). Also, the displacement of electrons can produce ' displaced ' atoms by breaking chemical bonds for a time long compared with the periods of thermal vibration. A striking example is the large irradiation effects observed in plastics (Sisman and Bopp 1951, Charlesby 1954). Such effects are akin to ionization induced chemical reactions and, although it is not possible to distinguish sharply between radiation chemistry and radiation damage, we shall only be concerned with the effects of displaced electrons in cases where they occur simultaneously with, or have an important bearing on the effects of the displaced atoms produced by direct momentum transfer. Likewise the introduction of foreign atoms is of secondary interest here.

1.2. *Historical*

Prior to the discovery of radioactivity, only relatively low energy radiation was available which, when capable of displacing atoms, has only a low penetrating power. Such radiation served, however, to reveal the phenomenon of sputtering (Grove 1852, Massey and Burhop 1952), the displacement of atoms from surfaces by the impact of ions, which is due essentially to the same mechanisms as those considered herein.

With the discovery of radioactivity and the isolation of sources of high specific activity, many bulk effects were observed (Rutherford, Chadwick and Ellis 1930), although most subsequent developments were concerned with their chemical aspects (Lind 1928, Hahn 1936). H. G. Wells (1909) dramatically visualized the possibility of mechanical failure of constructional materials under intense irradiation. Certain phenomena observed in minerals, the best known being the pleochroic haloes (Joly 1907) in mica and fluorspar, were accounted for by the bombardment due to radioactive elements.

As early as 1893, sufficient mineralogical data had been collected for Broegger (1893) to generalize and define *metamict* materials as originally crystalline materials which have in the course of time assumed the properties of amorphous substances. While retaining their original morphology, they have become less dense, assumed

conchoidal fracture (instead of cleavage) and become optically isotropic; some of them exhibit a sudden release of stored energy when they are heated (Berzelius 1815). All metamict crystals contain uranium or thorium, either as a normal constituent or as an impurity ($>\sim 0{\cdot}1\%$), and it is now clear that as originally suggested by Hamberg (1914), the metamict state is brought about by the bombardment of the α-particles and recoil nuclei produced in the radioactive decay of uranium and thorium. An admirable summary of present-day information on the metamict state has been given by Pabst (1952), and this, together with results published subsequently, makes it clear that an important part of the available observations on irradiation effects have been made on metamict minerals.

Since the development of nuclear reactors much progress has been made in understanding irradiation effects. Burton (1947), Slater (1951), Seitz (1949, 1952 a), Broom (1954), and Koehler and Seitz (1954) have discussed various aspects of the phenomena encountered. Dienes (1953 a b) has reviewed the experiments reported prior to March 1952. We endeavour in the present article, by using both the mineralogical data and the many results that have become available since 1952, to give a more detailed account than hitherto of the mechanisms involved.

1.3. *Relation between Theory and Experiment*

Very diverse phenomena are produced in solids by irradiation. Partly this is because the changes produced by irradiation are usually metastable; they depend critically on the structure of the solids, the nature of the interstitial sites available for the displaced atoms, the height of potential barriers, and suchlike detailed properties. Also some recovery usually occurs during or just after irradiation, and the magnitude of the effects is often very sensitive to temperature both of irradiation and of observation. Further, the subject is in a relatively early stage of development, so that at present a great many of the conclusions drawn await confirmation.

Consequently there is no simple set of rules, or basic theory to guide studies in this field. Instead, the possible processes taking place must be conjectured theoretically, and the experimental data in any instance must be analysed to find a combination of likely processes which fits the facts. These theoretical considerations of the number and grouping of the displaced atoms produced by irradiation, and of their recovery, are the main concern of § 2 and § 3.

§ 2 The Production of Displaced Atoms

2.1. *Energy Considerations*

2.1.1. *Displacement energy.*

From heats of sublimation and similar thermodynamic data, it is known that atoms are bound in solids with energies which are generally about 5 ev; it must be expected that more energy is needed to displace an atom from a lattice site within the solid. Firstly, more bonds are broken in displacing an atom inside a solid than in evaporating one from the surface; secondly, there must be a potential barrier between the lattice site and a stable interstitial site. Results of a detailed calculation for the case of copper have been given by Huntington (1954), who finds that the energy required depends on the direction (with respect to the

crystal axes) in which the displaced atom is knocked, and lies between 18 and 43 ev.

Experimental data from threshold energies for irradiation effects yield 31 ev for germanium (Klontz and Lark-Horovitz 1951), 25 ev for copper (Eggen and Laubenstein 1953), <12 ev for copper atoms in Cu_3Au (Dugdale 1955) and $\geqslant 25$ ev for iron in iron–copper alloy (Denney 1953). Threshold energies for sputtering are neither very reliable for application to irradiation damage, nor easy to obtain experimentally. They will, however, give an order of magnitude. Values ranging from 20 ev for Pb to 180 ev for Hf have recently been found in 23 metals by Werner (1954).

Since the displacement energy varies with direction of recoil, must vary from one material to another, and cannot be at all easily calculated, the easiest practical procedure is to follow Seitz (1949) in taking an isotropic value of 25 ev, and to modify this value only when directly indicated by experiment. This has generally been done in work so far published.

2.1.2. *Threshold radiation energies.*

With this working value of 25 ev for the displacement energy, the minimum radiation energy for the production of displaced atoms by the various types of radiation may readily be calculated. Results are presented in table 2.1. In the

Table 2.1. Threshold Radiation Energy for Displacements with $E_d=25$ ev

Atomic weight of stationary atoms	10	50	100	200
Neutrons, protons (ev)	76	325	638	1263
Electrons, γ-rays (Mev)	0·10	0·41	0·68	1·10
α-particles (ev)	31	91	169	325
Fission fragments (ev)	85	30	25	27

case of massive-particle radiation, the expression for the maximum energy E_{max} transferred by moving particle of mass M_1 and energy E to a stationary atom of mass M_2,

$$E_{max}=\Lambda E \qquad \ldots\ldots(2.1)$$

where

$$\Lambda=4M_1M_2/(M_1+M_2)^2 \qquad \ldots\ldots(2.1a)$$

has been used. For electron bombardment, relativistic effects are important, and we have used the expression

$$E_{max}=2E(E+2mc^2)/M_2c^2 \qquad \ldots\ldots(2.2)$$

where m is the electron mass, and c is the velocity of light. In the case of γ-rays, the threshold energy is determined by the energy of the photo- or Compton electrons, and therefore is the same roughly as the threshold for electrons. Direct interaction of γ-rays with the nucleus is rarely important except in the case of nuclear recoils from γ-emission, as in the Szilard–Chalmers process (McKay 1950).

2.2. *The Number of Displaced Atoms*

Atoms which are displaced by direct collisions with the bombarding radiation we call ' primary knock-ons ', to distinguish them from those atoms displaced as a consequence of the subsequent collisions of the primary knock-ons. In §§ 2.2.1–2.2.4, we deal generally with the collisions of moving atoms, and obtain expressions

both for the number of displaced atoms produced by a primary knock-on, and also for the cross sections for the production of primary knock-ons by bombarding ions. We use these expressions to calculate the total number of displaced atoms for certain cases, and in later sections, for the case of neutrons, electrons and γ-ray bombardment.

No accurate experimental determinations of numbers of displaced atoms are available, but some rough estimates provide comparisons between calculated and observed numbers, which we give in table 2.2. The calculated number is usually the greater of the two, probably because of the occurrence of recovery. In each

Table 2.2. Ratio of Calculated to Observed Numbers of Displaced Atoms

Material	Radiation	Calc. no. / Obs. no.	Assumptions
Copper	10 MeV deuterons	3	Electrical resistivity increased by $2\mu\Omega$ cm for 1% Frenkel defects
Silver	,, ,, ,,	4	
Gold	,, ,, ,,	4	
Germanium	1·5 MeV electrons	2	2 electrons trapped per Frenkel defect
Quartz	Fast neutrons	5	3% density change for 1% Frenkel defects
Boron nitride	Li and α-particles	>5	x-ray intensity changes detectable for $>$ 1% Frenkel defects.

case the experimental estimate depends on some sort of assumption relating the effects observed to the number of displaced atoms. These assumptions are briefly indicated in table 2.2; further details are given in § 5.

2.2.1. *Moving atoms: principles.*

Calculations for moving atoms have been made by Seitz (1949), Ozeroff (1949) and Brinkman (1954). Although the majority of Seitz's results are correct, his treatment is based on the use of the Born approximation, which is invalid for the collisions involved. Clear discussions of the ranges of validity of the different approximations have been given by Bohr (1948) and Williams (1945), and for all the collisions of interest herein, the classical approximation must be used. Ozeroff deals primarily with the case of atoms displaced by fission fragments. Brinkman, who corrects an error in Ozeroff's working, deals with the case of identical moving and stationary atoms, but does not explicitly calculate the number of displaced atoms. We base our discussion on the work of Knipp and Teller (1941) and Bohr (1948).

Our basic approximation is that the atoms are displaced by the Coulomb interaction between the nuclei, and that the effect of the screening electrons is to cut off this interaction sharply at a radius of about $a_0/Z^{1/3}$, where a_0 is the Bohr radius and Z is the atomic number. When the kinetic energy of the moving atom is less than the Coulomb interaction potential of the two nuclei separated by the radius of the screening electron cloud, the screen produces hard sphere collisions. The upper energy limit L_A for the validity of this approximation is (Bohr 1948)

$$L_A = 2E_R Z_1 Z_2 (Z_1^{2/3} + Z_2^{2/3})^{1/2} (M_1 + M_2)/M_2 \qquad \ldots\ldots (2.3)$$

where E_R is the Rydberg energy (13·60 ev), Z and M are atomic numbers and masses respectively, and the subscripts 1 and 2 refer to moving and stationary atoms respectively. At energies above L_A the effect of the screening electron

clouds is to cut off the Rutherford collisions at impact parameters of $a_0/(Z_1^{2/3}+Z_2^{2/3})^{1/2}$ and the minimum energy E^* that can be transferred is given by

$$E^*=4E_R^2Z_1^2Z_2^2(Z_1^{2/3}+Z_2^{2/3})M_1/M_2E. \qquad \ldots\ldots(2.4)$$

As long as E^* is more than the displacement energy E_d all Rutherford collisions displace atoms; but when the energy E of the moving atoms exceeds L_B where

$$L_B=4E^2_RZ_1^2Z_2^2(Z_1^{2/3}+Z_2^{2/3})M_1/M_2E_d \qquad \ldots\ldots(2.5)$$

only some of the Rutherford collisions do so. When $E\gg L_B$, it is easily shown that half the energy lost in Rutherford collisions is dissipated in lattice vibrations without displacing atoms.

The remaining interaction of importance is electronic excitation. Following Seitz (1949), we assume that energy loss from this source exceeds all others by a large factor ($\sim 10^3$) when $E>L_C$, and is zero when $E<L_C$; L_C is determined by the lowest excitation potential I_t of the valence electrons, and is roughly given in insulators by

$$L_C=(M_1/m)I_t/8 \qquad \ldots\ldots(2.6a)$$

where I_t is a well-defined value corresponding to the edge of the first main optical absorption band. In metals, where the valence electrons are free, electronic excitation losses occur at all energies. Ozeroff shows that at sufficiently low energies, these losses can be neglected in comparison with collision losses. No satisfactory criterion has been given for the energy above which electronic losses predominate, Seitz's expressions being invalid owing to his use of the Born approximation. Clearly, however, once the velocity of the moving atom is much less than that of electrons at the top of the Fermi distribution, only a small proportion of the electrons can be excited, and we take, by rough analogy with the case of insulators,

$$L_C=\frac{1}{16}\frac{M_1}{m}\epsilon_0\simeq\frac{1}{16}\frac{M_1}{m}\pi a_0^2E_R(3N_0\sqrt{\pi})^{2/3} \qquad \ldots\ldots(2.6b)$$

where ϵ_0 is the Fermi energy of the free electrons, and N_0 is the number of atoms per unit volume.

With heavy moving atoms, the assumption of a sharp cut-off to the ionization losses is very crude, owing to the large change in the effective charge of the moving atoms. Knipp and Teller (1941) show that ionization losses and collision losses are of about the same magnitude over quite a wide range of energy. Our values of L_C represent energy limits *below* which ionization losses may be neglected. We give some values of L_A, L_B and L_C in table 2.3.

Table 2.3. Energy Limits for Collision Processes

Substance	Moving Atom	Energy Limits (ev)		
		L_A	L_B	L_C
Diamond	Carbon	5×10^3	3×10^5	$1{\cdot}5\times10^4$a
Diamond	Deuteron	4×10^2	8×10^2	2×10^3a
Aluminium	Aluminium	3×10^4	9×10^6	$1{\cdot}7\times10^4$
Aluminium	Deuteron	1×10^3	2×10^3	$1{\cdot}3\times10^3$
KCl	Argon	6×10^4	4×10^7	7×10^4b
KCl	Deuteron	$1{\cdot}4\times10^3$	4×10^3	3×10^3b
Copper	Copper	2×10^5	4×10^8	5×10^4
Copper	Deuteron	3×10^3	8×10^3	$1{\cdot}6\times10^3$
Gold	Gold	2×10^6	4×10^{10}	$1{\cdot}2\times10^5$
Gold	Deuteron	1×10^4	4×10^4	$1{\cdot}2\times10^3$

The value of the first excitation level of the valence electrons is taken from: a, Moss (1952), b, Schneider and O'Bryan (1937).

2.2.2. *Moving atoms: hard sphere collisions.*

The essential feature of hard sphere collisions is that the probability of transferring a given energy ΔE, is independent of ΔE for $0 \leqslant \Delta E \leqslant \Lambda E$ and is zero outside this range. Consider a monatomic solid in which a primary knock-on of energy $\bar{E}$ has been produced. At its first collision the energy is shared by two atoms and at the second group of collisions the energy is shared by four atoms. Likewise, after the pth group of collisions the energy is shared by 2^p atoms and the number of atoms in an energy range dE is found to be

$$N_p(E)dE = 2^p(\ln(\bar{E}/E))^{p-1}\, dE/\bar{E}(p-1)! \qquad \ldots\ldots(2.7)$$

This result applies to free stationary atoms; but only those which receive more than the threshold energy E_d are displaced. Further, only an atom with energy greater than $2E_d$ can make a collision with a stationary atom, after which both have energy exceeding E_d. Thus we assume that atoms with energies between E_d and $2E_d$ are displaced but cannot themselves further increase the total number of displaced atoms. The number of such atoms produced by the pth group of collisions may be shown to be $E_d N_p(2E_d)$, so that the total number N_d of displaced atoms is

$$N_d = \sum_{p=1}^{p=\infty} E_d N_p(2E_d) = \bar{E}/2E_d. \qquad \ldots\ldots(2.8a)$$

This is valid only for $\bar{E} > 2E_d$; obviously for $0 \leqslant \bar{E} \leqslant 2E_d$,

$$N_d = 1. \qquad \ldots\ldots(2.8b)$$

Since (2.8*a*) is linear in $\bar{E}$ it is also correct when $\bar{E}$ represents the average energy of a group of primary knock-ons with energy greater than $2E_d$. It is easily shown that (2.8*a*) is also valid when the mass of the initial moving atom is different from that of the stationary atoms, provided $2E_d/\bar{E} \ll \Lambda$.

A simple argument due to T. M. Fry also leads to (2.8*a*). A net increase in the number of displaced atoms can be produced only when a moving atom with energy greater than $2E_d$ collides with a stationary atom. At such collisions, atoms appear in the energy range 0–$2E_d$ with uniform energy distribution and average energy E_d. The number of atoms which become involved in such collisions is therefore $\bar{E}/E_d$; consequently the number of displaced atoms is $\bar{E}/2E_d$.

The physical assumptions in the derivation of equation (2.8*a*) are such that it probably represents an overestimate of the number of displaced atoms. Seitz states that the formula

$$N_d = (E'/E_d)^{1/2} \qquad \ldots\ldots(2.9)$$

where E' is the average energy of *all* the primary knock-ons, is a good approximation when E' is a few times E_d. Comparison with equations (2.8*a*) and (2.8*b*) show that agreement is reasonable up to $\bar{E} \sim 10E_d$. Ozeroff gives an approximate treatment based on the Thomas–Fermi model, which allows for the electrostatic interaction of the screening electron clouds. His results, which cannot be presented in a simple analytical form, give values of N_d some 30% less than equation (2.8*a*). The calculations contain the error pointed out by Brinkman and some imperfectly justified assumptions; the results can therefore be used only as a rough guide to the accuracy of equation (2.8*a*).

2.2.3. *Moving atoms: Rutherford collisions.*

When $E>L_B$ the cross section σ_p for a Rutherford collision in which energy greater than E_d is transferred is readily found, by applying the ordinary theory for unscreened nuclei (Bohr 1948), to be

$$\sigma_p=4M_1Z_1^2Z_2^2E_R^2(1-E_d/\Lambda E)\pi a_0^2/M_2E_dE. \qquad \ldots\ldots(2.10)$$

The symbols are defined in § 2.2.1; πa_0^2 has the value 8.8×10^{-17}cm². The term $(1-E_d/\Lambda E)$ is very close to unity for all cases of interest, and may be omitted, so that the cross section increases as the energy of the moving atom decreases. It follows from equation (2.10) that the number of primary knock-ons varies inversely as the square of their energy, and their mean energy $\bar{E}$ is given by

$$\bar{E}=E_d\ln(\Lambda E/E_d). \qquad \ldots\ldots(2.11)$$

As Seitz points out, the logarithmic term is insensitive to the exact values of Λ and E, and $\bar{E}$ is between 3 and 15 times E_d for most practical cases. At such energies the hard sphere results may be used to calculate the number of atoms displaced by the primary knock-ons. From equations (2.8*a*) and (2.8*b*) the average *total* number N_d of displaced atoms produced for every primary knock-on is

$$N_d=\tfrac{1}{2}\{1+\ln(\Lambda E/2E_d)\}. \qquad \ldots\ldots(2.12)$$

These equations may be applied directly to such cases as the cyclotron irradiation of thin foils; for copper irradiated with 10 Mev deuterons, we obtain $\sigma_p=6900$ barns, $\bar{E}=2700$ ev and $N_d=5{\cdot}8$.

When $L_A<E<L_B$, as in the case of fission fragments and energetic primary knock-ons, all the Rutherford collisions produce displaced atoms, and σ_p is approximately constant and equal to $\pi a_0^2/(Z_1^{2/3}+Z_2^{2/3})$. The average energy of the primary knock-ons and N_d can be estimated by replacing E_d by E^* in equation (2.11).

Both Brinkman and Ozeroff refine our approximation of a sharp cut-off of the nuclear interaction, by considering the energy transfer between two Thomas–Fermi atoms with rigid charge distributions. Brinkman gives an expression of σ_p for the case $Z_1=Z_2$ valid at all energies which confirms the condition $E>L_B$ for equation (2.10), and which shows the expected slow variation of σ_p with E when $E<L_B$. Ozeroff's approximate results show that the nuclear Rutherford collisions are the predominant source of energy loss, down to energies at which the hard sphere interaction becomes dominant. Both these workers use an approximate form of the Thomas–Fermi potential, and take the radius of the screening electron cloud as $2{\cdot}09a_0/Z_2^{1/3}$; Knipp and Teller give the value $0{\cdot}85a_0/Z_2^{1/3}$ (assuming $Z_2>Z_1$), and Bohr gives $a_0/(Z_1^{2/3}+Z_2^{2/3})^{1/2}$. We have used Bohr's value.

2.2.4. *Moving atoms slowing down to rest.*

We have already shown that the total number of displaced atoms produced when an atom of energy E slows down to rest is $E/2E_d$ when $E<L_A$. Provided $E<L_C$ this result also holds for $L_A<E<L_B$, since the predominant energy loss in this range is then due to Rutherford collisions which produce primary knock-ons (see also Snyder and Neufeld 1954). When $E>L_B$ only some of the energy is expended in this way, and the result no longer holds; for the case $E\gg L_B$ the total number of displaced atoms is approximately $E/4E_d$.

Usually, however, when $E>L_B$, E is also greater than L_C and electronic

excitation losses have to be taken into account. In the high energy case of $E \gg L_C$, $E \gg L_B$, these losses, which predominate, are given by

$$-(dE/dx)=2\pi(qZ_1)^2e^4Z_2N_0(M_1/mE)\ln(4Em/M_1I) \quad \ldots\ldots(2.13)$$

where e is the electronic charge, N_0 is the number of atoms per unit volume, I is a mean ionization potential of the stationary atoms, and qZ_1 is the effective charge of the moving atom. Full discussions of equation (2.13) and the appropriate values of I and q are given by Bethe and Ashkin (1953) and by Allison and Warshaw (1953). Combining equations (2.13), (2.10) and (2.12), it is found that in losing energy ΔE, a moving atom displaces ΔN_t atoms where

$$\Delta N_t=P\Delta E/E_d \quad \ldots\ldots(2.14)$$

and

$$P=mZ_2\{1+\ln(\Delta E/2E_d)\}/4M_2q^2\ln(4Em/M_1I). \quad \ldots\ldots(2.14a)$$

When $q \simeq 1$, P is usually between 10^{-3} and 10^{-4}.

For very light moving atoms such as α-particles produced by radioactive disintegrations, q and I are constant over a large proportion of the energy range, and P is so insensitive to E that a mean value can be used without serious error. Furthermore, the energy range L_B–L_C is small, and equation (2.14) can be applied with a constant mean value of P to all energies in excess of L_C. Thus the total number N_t of displaced atoms produced by a very light moving atom of energy E_0 in slowing down to rest is

$$N_t=\{P(E_0-L_C)+bL_C\}/E_d \quad \ldots\ldots(2.15)$$

where P is given by equation (2.14a), L_C is given by equation (2.6), and b is $\frac{1}{2}$ if $L_B > L_C$ and $\frac{1}{4}$ if $L_B \ll L_C$. Essentially the same result is obtained by Seitz. In cases where the term bL_C predominates, a great deal of uncertainty is introduced by the procedure for obtaining L_C, which is hardly more than qualitative.

The case of heavy moving atoms with $E > L_C$ is more complex because ionization losses are predominantly governed by variations in the effective charge. Also L_B is usually so large that equation (2.10) no longer applies over a large energy range, if at all. This case is dealt with in detail by Ozeroff. Essentially, equation (2.13) is replaced by a semi-empirical expression, E_d is replaced by E^* in equation (2.11) and a series of averages is taken to circumvent the mathematical complexities. This procedure is essential in the case of fission fragments. For the case of fast-moving primary knock-ons, the initial energy is not often greatly in excess of L_C, and it is often sufficient to suppose that *all* the excess energy is expended in ionization losses, so that the number of atoms displaced is $L_C/2E_d$.

2.2.5. *Fast neutrons.*

The fast neutrons used in irradiation work are usually produced by fission. Watt (1952) gives an energy spectrum, having a maximum at 0·7 MeV and a mean energy of 2 MeV, for fission neutrons from both ^{235}U and ^{239}Pu.

The collisions of the neutrons are of the simple hard-sphere type, and so the recoil energy spectrum $N(E)dE$, of the struck atoms is given by

$$N(E)dE=dE/E_{max}(0 \leqslant E \leqslant E_{max}) \quad \ldots\ldots(2.16)$$

where E_{max} is given by equation (2.1). Only the negligible proportion E_d/E_{max} of the struck atoms are not displaced, so that σ_p the cross section for producing a

primary knock-on is the total neutron cross section. For the heavy elements, E_{max} does not usually exceed the ionization limit L_C. Applying the results of the previous sections, the total number N_d displaced on the average for each fast neutron collision is given by

$$N_d = E_{max}/4E_d. \qquad \ldots\ldots(2.17)$$

For lighter elements E_{max} generally exceeds L_C, but except for very light elements, the primary knock-ons effectively lose *all* the excess energy in ionization, and in this case

$$N_d = (2 - L_C/E_{max})L_C/4E_d. \qquad \ldots\ldots(2.18)$$

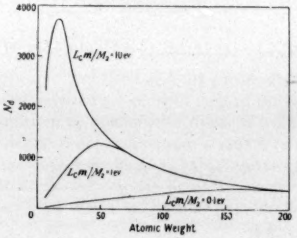


Figure 2.1. The average number of displaced atoms per primary knock-on produced by 2 MeV neutron bombardment as a function of atomic weight.

Values of N_d for 2 MeV neutrons are presented at figure 2.1, for various values of the parameter $L_C m/M_2$. This parameter is not a sharply varying function of M_2, and is of the order of 1 ev in many materials. The shape of the curves shows that 2 MeV neutrons produce the maximum number of displacements for moderately heavy elements. Heavy nuclei take up less recoil energy; lighter nuclei receive more energy but lose a greater proportion in ionization (Burton 1947).

Ozeroff (1949) has applied his more complex expressions, referred to in § 2.2.3, to the number of atoms displaced by fast neutrons, and both he and Seitz give some values for the case of a fast neutron slowing down to rest.

2.2.6. *Electrons.*

Electrons displace atoms by Coulomb interaction with the nuclei. They invariably have to penetrate the K electron shells to do so, and therefore expressions derived on the basis of an unscreened Coulomb field can safely be used. However, we cannot use equation (2.10) because the electrons normally have relativistic velocities. The relativistic expression of Mott (1929, 1932) for the differential scattering cross section forms the basis of the calculation. This expression cannot be put into a rigorous explicit form. Dugdale and Green (1954), using the α^2 approximation of McKinley and Feshbach (1948), obtain

$$\sigma_p = \frac{4Z_2^2 E_R^2 E_{max}}{m^2 c^4 E_d^2} \frac{1-\beta^2}{\beta^4}$$

$$\times \left[1 + 2\pi\alpha\beta\left(\frac{E_d}{E_{max}}\right)^{1/2} - \frac{E_d}{E_{max}}\left\{1 + 2\pi\alpha\beta + (\beta^2 + \pi\alpha\beta)\ln\left(\frac{E_{max}}{E_d}\right)\right\}\right] \pi a_0^2 \quad \ldots\ldots(2.19)$$

where E_{max} is given by equation (2.2), β is the electron velocity divided by the velocity of light, and $\alpha = Z_2/137$. McKinley and Feshbach estimate that their α^2 approximation is satisfactory when $Z \lesssim 27$ and $\beta \simeq 1$; for the calculation of σ_p, equation (2.19) is probably adequate well outside these limits. Klontz (1952) and Denney (1954 a) give essentially similar expressions based on Mott's own approximation; there are some discrepancies which can be rectified with the aid of Bethe and Ashkin (1953, p. 280).

The variation of σ_p with electron energies is shown in typical cases in figure 2.2. For small values of β, equation (2.19) reduces to (2.10). For electrons, the variation of σ_p with energy is predominantly due to the term $1-E_d/\Delta E$, since $\Delta E \sim E_d$, whereas for ions, this term is negligible. Thus in contrast with ions, σ_p for electrons increases with energy in the range of practical interest. Morever, the

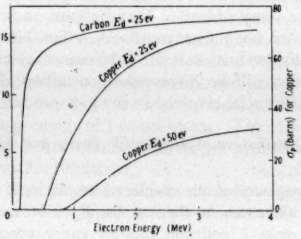


Figure 2.2. Cross section for producing primary knock-ons by electron bombardment as a function of energy.

relativistic factors usually intervene before any decrease similar to that of equation (2.10) can be observed. At high energies $M_2c^2 \gg E \gg mc^2$, $E_{max} \gg E_d$ the cross section saturates at a value given by

$$\sigma_p = \frac{8E_R^2 Z_2^2}{M_2 c^2 E_d} \pi a_0^2. \qquad \ldots\ldots(2.20)$$

To a first approximation, the number of primary knock-ons varies as the inverse square of their energy, and their mean energy is given by the equivalent of equation (2.11),

$$\bar{E} = \frac{E_d E_{max}}{E_{max} - E_d} \ln\left(\frac{E_{max}}{E_d}\right). \qquad \ldots\ldots(2.21)$$

In many cases, E_{max} is no more than a few times E_d—for instance in copper it is 68 ev for 1 Mev electrons—and the total number of displaced atoms is not greatly in excess of the number of primary knock-ons.

Denney (1954 a) has shown that recoil energy taken up by bremsstrahlung represents only a small correction, and has also integrated σ_p over the range of the electrons.

2.2.7. *Gamma rays.*

The production of displaced atoms by γ-rays has received very little attention so far. However, Dugdale (1955) has shown that measurable effects can be produced. Most of the results required for obtaining cross sections are given by Bethe and Ashkin (1953). For γ-ray energies of about 2 Mev and materials with

$Z_2 \lesssim 50$ the predominant interaction is the Compton effect. Atoms are displaced by the recoil Compton electrons. For the case of copper irradiated with 5 MeV γ-rays, σ_p is about 0·1 barn.

For heavy elements the direct displacement of the atoms by recoil from the photoelectric effect is predominant. The cross section σ_K for photoelectric absorption (for $E \gg mc^2$) is approximately

$$\sigma_K = Z_2\alpha^4(mc^2/E)\exp\{-\pi\alpha+2\alpha^2(1-\ln\alpha)\}\text{barns.} \quad \ldots\ldots(2.22)$$

$\alpha = Z_2/(137)$. For 5 MeV γ-rays, this gives cross sections of the order of 10^{-24} cm^2 for the very heavy elements. Not every photoelectric absorption will displace the atom concerned, since the distribution of the emitted electrons with respect to the incoming γ-ray has to be taken account of, and σ_K is likely to be an overestimate for σ_p.

None of the other γ-ray processes would appear to be as important at these energies as the two we have discussed. At rather higher energies ($\gtrsim$10MeV, depending on Z_2), pair production is the predominant γ-ray process, and it must be expected that atoms will be displaced by recoiling electrons and positrons, and by the recoil of the nuclei involved in the pair-production.

2.3. *Distribution of Displaced Atoms and Vacancies*

2.3.1. *Collision models.*

In all cases, the *majority* of the displaced atoms leave their lattice sites with energies of no more than two or three times E_d. At these energies, a moving displaced atom will make a collision roughly every mean interatomic distance travelled, and will lose half its energy on the average at each collision. If we suppose that it becomes trapped in an interstitial site when it is slowed down to 1 eV, a displaced atom will usually have made no more than six or seven collisions before being trapped, and will thus be only a few interatomic distances away from the vacancy left behind.

The simplest case is provided by electron or γ-ray bombardment when the primary knock-ons have insufficient energy to displace further atoms. The radiation damage consists of a random distribution of Frenkel defects in which the distance between the interstitial and the vacancy can rarely be more than four or five interatomic distances. When the electrons can impart more than $2E_d$ to the primary knock-ons, these will displace further atoms, and some of the single Frenkel defects will become groups of two or more interstitial atoms together with the corresponding vacancies, all within a few interatomic distances of each other.

In the case of a fast-moving atom, along the first part of the track where $E > L_B$ and equations (2.10) and (2.11) are valid, the damage consists of groups of several vacancies and interstitial atoms, widely separated along the track of the ion. As the atom slows down, the cross section for producing primary knock-ons increases and the groups become more frequent; when $E \ll L_B$ a primary knock-on is produced for roughly every interatomic distance travelled (Brinkman 1954), and so the concentration of interstitials and vacancies is very high at the end of the track. This feature of irradiation damage has been found experimentally by Brattain and Pearson (1950) and by Meyer (1954). It is also a characteristic feature of the pleochroic haloes observed in mica, where the best defined haloes are shells of metamict material (Hurley and Fairbairn 1953).

Fast neutron bombardment must usually produce the highly concentrated type of damage last mentioned. When the energy of a knock-on atom is sufficiently low, the mean free path between collisions in which a large fraction of the energy is lost approaches the interatomic distance. Under these conditions, the collision models suggest that the damage will consist of a region containing a very high concentration of vacancies, surrounded by a region with a high concentration of interstitial atoms.

2.3.2. ' *Displacement spikes* ' *and* ' *thermal spikes* '.

Brinkman (1954) has suggested that with the highly concentrated damage obtained with fast neutron bombardment of heavy elements, it is pointless to speak of individual collisions and has proposed that the volume affected is melted for a short period, and then solidifies, in the main with the same crystallographic orientation as the parent material. Such volumes he calls *displacement spikes*. The size of the displacement spike is determined as follows. As a moving atom slows down, the cross section for displacement collisions increases; at the point where one displacement collision occurs for every interatomic distance travelled, *all* the remaining energy of the moving atom is distributed in a displacement spike with an average energy of 1 ev per atom. The average size of a displacement spike produced by 2 Mev neutrons in copper is thus estimated to be 75 Å in diameter, and contains 2×10^4 atoms.

Brinkman suggests that the main defects introduced by such spikes would be small dislocation loops, and a small number of interstitial atoms and vacancies quenched in on solidification.

A rather similar model—the *thermal spike* model—has been proposed by Seitz (1949), in which a moving particle heats up the material surrounding its track through the solid for times of the order of 10^{-12} second. As in the case of displacement spikes, the consequences of the rapid heating and cooling cannot be predicted at all clearly (Koehler and Seitz 1954). Experimentally, it is observed that the ordered alloy Cu_3Au is disordered by neutron bombardment more rapidly than would be expected simply from the number of atoms displaced, and this is consistent with the displacement spike model.

The practical validity of these models depends on the limits of volume and time to which the macroscopic concepts of heat can be applied. The frequencies of atomic vibrations are in the range 10^{12}–10^{13} sec^{-1}. During a time of 10^{-12} second, a disturbance to the lattice is propagated by elastic waves for distances of between 10 and 50 Å, which is also about the mean free path of phonons in insulators at room temperature (Kittel 1953, p. 82). Consequently, in insulators, it is unsafe to apply the macroscopic concepts of heat for times shorter than 10^{-12} second, or to volumes with linear dimensions much less than 50 Å. In metals, where free electrons play a dominant role in the transport of energy, elastic disturbances are mainly converted to disturbances of the free electrons in times of about 10^{-13} second (Kittel 1953, p. 243). The mean free path associated with the transport of energy by free electrons is of the order of 100 Å or more, and so a rather larger volume is probably necessary in metals than in insulators before the macroscopic concepts of heat can be applied.

With these rather rough criteria, it can be seen, for instance, that the thermal spike model of the damage produced by a fission fragment moving in uranium is probably valid. A 100 Mev fragment has a range of about 4×10^{-4} cm (Ozeroff 1949);

it is easily shown that a cylindrical volume of 100 Å radius enclosing the track is heated to about 4000°c by the dissipation of 100 MEV. Classical heat-conduction expressions show that the temperature on the track falls to half this value in about 10^{-11} second. In contrast, similar calculations suggest that the damage produced by electrons, γ-rays and energetic ($E>L_B$) light ions is probably best regarded as being produced by individual collisions.

The damage produced by neutrons is usually of an intermediate character and the criteria are too ill-defined to provide a decision. Experimentally there is no doubt that fast neutrons produce interstitial atoms, but at present it is not clear whether these are best regarded as being produced by rapid cooling of small volumes, or by individual collisions. It is quite possible that both models should be used simultaneously, for the primary knock-ons have all energies from E_d to E_{max}, and the low energy knock-ons are almost certainly regarded best on the collision model. For general pile irradiations, because of the shape of the energy distribution of the fast neutrons (equation (4.1)) the majority of primary knock-ons have relatively low energies. Evidence that fast neutron damage can be regarded as a mixture of isolated interstitials and relatively larger volumes of concentrated damage is provided by Berman (1951) and Klemens (1951) in studies on quartz.

The same difficulty of approach has been found in attempts to account quantitatively for sputtering, and Bradley (1954) concludes that no single model so far used gives an adequate account under all conditions. He finds, however, no evidence in favour of a ' hot-spot ' model for ions of energies up to 1800 ev, incident on alkali metals. The most explicit application of the ' hot-spot ' model, that of Townes (1944) is open to the objection that the macroscopic concepts of heat cannot be applied to volumes containing only a few atoms for times of about 10^{-16} second. Keywell (1952) has applied a hard-sphere collision model to the sputtering caused by argon ions with energies of up to 5000 ev in silver and other metals, and obtains quite a good agreement with his observations, assuming that one silver atom is sputtered for every collision made by the argon ion in slowing down to 39 ev.

2.4. *Effects in Compounds*

2.4.1. *Replacement collisions.*

Some of the collisions between moving atoms and stationary atoms at lattice sites will displace the initially stationary atom, but leave the original moving atom with insufficient energy to escape from the vacancy thus created. Such collisions may be called replacement collisions, since they change the identity of the displaced atoms but not the total number. Replacement collisions are clearly of little consequence in monatomic solids, but in compounds they may be important.

It has been proposed (Kinchin and Pease 1955) that a replacement collision takes place when a moving atom is left with energy that is both less than E_d and less than the energy given to the stationary atom. It is further assumed that a replacement collision can only occur when the energy imparted to the stationary atom exceeds some threshold value E_p. An approximate calculation of N_p, the number of replacement collisions, for the case of neutron bombardment, based on hard-sphere collisions and assuming all the atoms have the same mass, leads to

$$N_p=0{\cdot}5N_d[1+\ln(E_d/E_p)/\ln(4/3)] \qquad \ldots\ldots(2.23)$$

where N_d, the number of displaced atoms, is given by equation (2.17). Since E_o is likely to be appreciably less than E_d, the replacement collisions may well outnumber the displacement ones. This is particularly so in the case of ordered alloys, where E_o may be not much greater than the activation energy for interstitialcy diffusion which has been estimated to be about 0·25 ev (Huntington 1953) in copper.

If the masses of the atoms in a compound are very different, the number of replacements is almost certainly very much smaller than suggested by equation (2.23), for, as can be seen from equation (2.1), it is impossible for a moving atom of mass m_1, to give up more than half its energy to an atom of mass M_2, if A is less than a half. It is easily shown that, for replacements to occur by the mechanism proposed, $1/5{\cdot}83 < M_1/M_2 < 5{\cdot}83$.

2.4.2. *Disordering.*

In any compound, both replacement collisions and partial or complete recombinations of the Frenkel defects may produce atoms in lattice sites of the wrong kind. Such disordering is well known in the case of ordered alloys such as Cu_3Au and Ni_3Mn. There is little evidence available for other types of material. In the case of boron nitride, however, no disordering was detected even at very high doses (Pease 1954). Even in such a case, the irradiation must produce some defects in which a displaced atom is either in or very close to a vacant lattice site of the wrong kind. Such defects may be important in minerals, since these often contain many different kinds of atoms.

§ 3. Recovery

In most cases, on raising the temperature of a specimen after irradiation, the physical properties revert to their initial values. In addition, energy is released. In a few cases, after very heavy irradiation, annealing at high temperatures may cause a material to recrystallize in a new phase with a structure different from the original, as in the cases of thorite (Pabst 1952) and some forms of quartz (Wittels and Sherrill 1954). Usually one must expect some sort of recovery to take place during irradiations, at however low a temperature these are carried out.

Recovery produced by heating usually takes place in several stages, with different activation energies in different ranges of temperature, and the course of a given recovery process can usually be represented by the equation:

$$dn/dt = -cn^{\gamma}\exp(-\epsilon/kT) \qquad \ldots\ldots(3.1)$$

where n is the number of defects taking part in the recovery process, ϵ is the activation energy, c is a constant and γ is often termed the 'order of reaction' by analogy with gaseous chemical reactions. This terminology is somewhat misleading, since a number of recovery processes governed by the diffusion of lattice defects give apparently high orders of reaction which are not related to the number of constituents involved, as is the case in chemical reactions.

3.1. *Recombination of Interstitials with Vacancies*

3.1.1. *Close pairs of interstitials and vacancies.*

The recombination of interstitial–vacancy pairs is discussed by Fletcher and Brown (1953). They assumed that vacancies are the more mobile defects, and

applied diffusion equations to the random migration of vacancies to spherical sinks representing interstitials. If $n(0)$ vacancies start out at a distance b from a sink of radius a in an infinite isotropic three-dimensional medium, the number, $n(t)$, remaining after a time t is given by

$$n(t)=n(0)\left\{1-\frac{a}{b}+\frac{a}{b}\,\mathrm{erf}\left(\frac{b-a}{2(Dt)^{1/2}}\right)\right\} \qquad \ldots\ldots(3.2)$$

where D is the diffusion coefficient for vacancies. A fraction $1-a/b$ of the vacancies is not caught by the sink, but is 'liberated' to diffuse indefinitely throughout the medium. In the early stages of 'liberation', the lattice cannot be regarded as a continuum, and Fletcher and Brown therefore used numerical solutions, applicable to the diamond lattice.

For substances with a layer structure where the diffusion of interstitial atoms (assumed to be more mobile than vacancies throughout the remainder of § 3) can only take place in two dimensions; the spherical sink of the three dimensional problem is replaced by a circular sink. The solution is:

$$n(t)=n(0)\frac{2}{\pi}\int_0^\infty \frac{\exp(-\gamma^2 Dt)}{y}\left\{\frac{Y_0(by)J_0(ay)-Y_0(ay)J_0(by)}{J_0^2(ay)+Y_0^2(ay)}\right\}dy \quad \ldots.(3.3)$$

where J_0 and Y_0 are zero order Bessel functions. This equation has been evaluated by K. T. Spinney (unpublished) and some results are shown in figure 3.1. In

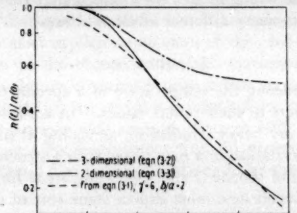


Figure 3.1. Annealing of close pairs of interstitials and vacancies.

this case there is no 'liberation' of interstitials. Equation (3.1), with $\gamma=6$, is a good approximation to equation (3.3).

3.1.2. *Random distribution of interstitials and vacancies.*

The combination of equal numbers of randomly distributed interstitials and vacancies follows the kinetics of a true second order reaction, i.e.

$$dn/dt=cn^2\exp(-\epsilon/kT) \qquad \ldots\ldots(3.4)$$

where ϵ is the activation energy for interstitial migration.

In the case of a material bombarded with electrons with energy just above the threshold energy, so that isolated Frenkel defects are produced, two stages of recovery would be expected, the first governed by equation (3.2) or (3.3) and the second by equation (3.4), with the same activation energy. The situation is complicated by the interaction between interstitials and vacancies, discussed by

Marx *et al.* (1952) and Overhauser (1953). They conclude that the lattice strain near a vacancy will reduce the activation energy for migration of a nearby interstitial and diffusion will no longer be isotropic. For short irradiations, however, the activation energy should rise during the liberation process to a value corresponding to the activation energy for migration in the undistorted lattice and the second stage should follow equation (3.4). Dienes (1953 c) has emphasized the pronounced effect of relatively small changes of activation energy on the apparent order of reaction.

3.2. *Competing Processes*

The recombination discussed above is too simple to explain all the observed effects. Marx *et al.* (1952) have outlined a number of other possible processes. During the annealing some interstitials may escape to dislocations or grain boundaries, thus affecting the kinetics of the interstitial–vacancy recombination. Also, after the disappearance of all the interstitials the excess of vacancies diffuse to dislocations and grain boundaries with higher activation energy. Approximate solutions for the relevant diffusion problems have been given by Fletcher and Brown (1953) and Brinkman *et al.* (1954). Groups of interstitials and of vacancies, which only dissociate with high activation energies, may also be formed.

Thus, although experimental studies of the kinetics of recovery may afford useful confirmatory evidence, they are insufficient to permit unambiguous identification of the processes. Better evidence about the recovery processes is secured by experimental measurement of well-defined activation energies, constant during comparatively large changes of the physical property measured. These can then be compared with estimates obtained from other experimental data (from cold work, quenching and self-diffusion experiments, for example) and from theoretical calculations.

3.3. *Accumulation of Damage*

We now consider the case of simultaneous irradiation and annealing. Irradiation generally produces groups of displaced atoms and vacancies. Provided that the interstitials diffuse for distances small compared with the distances between the groups, each group will anneal independently in the same way as the others. Hence, if dx displaced atoms are produced instantaneously by the irradiation, after a time t at a temperature T, only $dx\ \mathrm{F}(T,t)$ will be left, where $\mathrm{F}(T,t)$ is a function, independent of x, which may be determined experimentally by annealing after irradiation at very low temperatures. The total number of displaced atoms after a time t, during an irradiation in which displaced atoms are produced at the rate dx/dt' is then:

$$n(t)=\int_0^t\left(\frac{dx}{dt'}\right)\mathrm{F}(T,t-t')dt'. \qquad \ldots\ldots(3.5)$$

When dx/dt is constant throughout the irradiation, equation (3.5) shows that the number of displaced atoms present after a given *time* of irradiation is proportional to the flux of bombarding particles and also that, for different fluxes, the number is not determined solely by the total dose.

If the annealing of groups can be represented empirically by equation (3.1) with $\gamma>1$ and constant activation energy, $\mathrm{F}(T,t)$ becomes:

$$[1+c't\exp(-\epsilon/kT)]^{1/(1-\gamma)}$$

where c' is a new constant. For $\gamma>2$, the number of displaced atoms present after a time long compared with $(1/c')\exp(\epsilon/kT)$ is given approximately by

$$n(t)\simeq\frac{d\gamma-1}{d\gamma-2}\left\{\frac{1}{c'}\exp\left(\frac{\epsilon}{kT}\right)\right\}^{1/(\gamma-1)}t^{(\gamma-2)/(\gamma-1)}. \qquad \ldots\ldots(3.6)$$

For large values of γ the damage is almost proportional to x and the effect of simultaneous recovery is simply to reduce the apparent rate of production of displacements by the factor $\{(1/c')\exp(\epsilon/kT)\}^{1/(\gamma-1)}$. This temperature dependence of the rate of accumulation of damage, which would not be expected for low doses if the interstitials and vacancies were randomly distributed, is often observed (e.g. Marx *et al.* 1952).

3.4. *Saturation*

There is no term in equation (3.6) indicative of any saturation. This is because interaction between different groups of displaced atoms has been neglected. In fact, as irradiation proceeds the groups will eventually overlap and, at low temperatures when the interstitials are immobile, this will give rise to an exponential saturation of the number of displaced atoms. At higher temperatures, the approach to saturation will be governed by second order recovery.

Rather similar ideas emerge from a different approach used by Pease (1954)

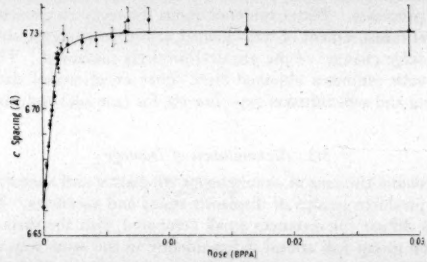


Figure 3.2. Saturation of change of lattice spacing in boron nitride. Boron nitride is a layer structure, and the c spacing is twice the interlayer spacing. The dose unit is the number of bombarding particles (α-particle and Li nuclei) per atom (Pease 1954).

in a study of saturation effects in boron nitride (figure 3.2). This used the approximation that interstitials and vacancies are stable if their separation exceeds a critical distance and that they instantly combine for smaller separations. Each interstitial atom is thus surrounded by Q lattice sites to which it can move immediately when one of them becomes vacant. Statistical arguments lead to an expression relating the concentration, y, of interstitial atoms, to the average number, p, of times each atom has been knocked into an interstitial site, which for large Q reduces to

$$y=(1/Q)\ln[2-\exp(-pQ)]. \qquad \ldots\ldots(3.7)$$

This expression gives $y=p$ at low doses and a saturation concentration of $y=(\ln 2)/Q$ setting in at a dose corresponding to $p\simeq 1/Q$. The effects of the temperature of

irradiation are introduced into this model by supposing that Q is temperature dependent, and that p is thus also dependent on temperature because of the marked grouping of interstitials and vacancies produced by the bombardment.

Even at the lowest temperatures, Q can hardly be less than the number of lattice sites immediately adjacent to an interstitial site, and thus the maximum concentration of interstitials which a crystalline material can support is unlikely to be much greater than 5%.

3.5. *Radiation Annealing*

Effects are sometimes observed which suggest that the recovery processes are hastened by the irradiation itself. Thus Cooper *et al.* (1954) observed that the rate of increase of resistivity of copper, silver and gold at 12°K decreased as irradiation proceeded, while no annealing was observed at this temperature after irradiation.

One obvious mechanism for this is the local heating produced by the passage of bombarding radiation. It seems likely, however (Koehler and Seitz 1954), that in most cases the duration of the heating is too brief to produce appreciable recovery. Koehler and Seitz suggest that the excitation of electrons facilitates the movements of interstitials or vacancies. It is also possible that collisions move the vacancies and interstitials. The threshold energy for such movements is much less than the 25 ev required for displacement and consequently the probability of moving an interstitial or a vacancy is much larger than the probability of forming it in the first place. Thus the irradiation may induce movements which are large compared with thermally activated movements, and produce a recovery which would not occur in the absence of radiation. This might appreciably reduce the saturation concentration of displaced atoms, referred to in § 3.4. This saturation may be important in limiting the concentration of displaced atoms in the groups produced by fast neutron bombardment (§ 2.3).

§ 4. Experimental Irradiation Techniques

4.1. *Reactor Irradiation*

4.1.1. *Fast neutrons.*

The radiation inside a reactor generally consists of fast and thermal neutrons, γ-rays and a small number of β-rays. The proportions of these radiations vary not only between one reactor and another, but also from one part of a reactor to another. In the majority of cases, however, the overwhelming majority of displaced atoms are produced by fast neutron bombardment.

Between the fuel rods of a graphite-moderated reactor the fast neutron energy spectrum is given roughly by the equation

$$F(E)dE = \frac{kdE}{E} \qquad \ldots\ldots(4.1)$$

where $F(E)\,dE$ is a flux and k is a constant (e.g. Taylor 1952). In many cases the fast flux is approximately equal to the thermal flux. In addition, the fission neutron spectrum is superimposed in a proportion which depends on the exact position with respect to the fuel rods. Inside a hollow uranium cylinder, the damage is predominantly due to fission neutrons. Such hollow cylinders provide

regions of high and uniform fission-neutron flux of approximately the same total magnitude as the general flux of thermal neutrons in the reactor core. The maximum thermal flux obtainable in BEPO (Harwell) is 2×10^{12} neutrons cm^{-2} sec^{-1}, a figure which is typical of most fair-sized graphite-moderated natural-uranium reactors. Higher fluxes are obtainable in heavy-water-moderated reactors such as the NRX pile at Chalk River. The recently constructed Materials Testing Reactor (Huffman 1954) has the highest fluxes of neutrons so far available, approximately 2×10^{14} neutrons cm^{-2} sec^{-1}.

Both the energy spectrum and the total integrated flux are needed to specify an irradiation dose. It is usual to choose a fixed position in a reactor and measure doses in terms of integrated thermal flux at that point. Fast neutron dose can then be estimated from theoretical or experimental knowledge of the ratio of fast to thermal flux. Neither the theoretical nor the experimental determinations of these ratios are at all easy or accurate (Barschall *et al.* 1952).

Fast neutrons produce primary knock-ons uniformly throughout large specimens, and the size of the specimens is primarily limited by practical considerations such as the size of experimental holes and the difficulty of handling radioactive materials. Temperature control of irradiations is somewhat troublesome at temperatures other than that of the reactor; controlled temperatures as low as $-150°$c have been reported (Blewitt and Coltman 1954, McReynolds *et al.* 1954).

The presence of γ-radiation and nuclear transmutations induced by thermal neutrons may produce ionization effects and impurity effects. The first of these cannot be avoided by any method of irradiation so far used; the second can be serious in materials where the transmutation cross section greatly exceeds the scattering cross section. Inside hollow uranium cylinders the high ratio of fast to slow flux reduces transmutation effects considerably; for the worst cases the thermal flux can be further reduced by cadmium shielding (Cook and Cushing 1953).

4.1.2. *Thermal neutron reactions.*

Thermal neutron reactions may be turned to good use in a few cases where the reactions produce massive particles. The three most useful reactions of this type are:

$^{6}Li(n,\alpha)^{3}H+4{\cdot}8$ Mev, natural cross section 70 barns.
$^{10}B(n,\alpha)^{7}Li+2{\cdot}4$ Mev, natural cross section 750 barns.
$^{235}U(n,f)$ fission fragments $+160$ Mev, natural cross section 3·92 barns.

The cross sections are for 0·025 ev neutrons, and the energies quoted are the average kinetic energy shared by the massive particles and recoil nuclei. When the fast and thermal neutron fluxes are equal, these reactions all produce displaced atoms at some ten to a hundred times the rate produced by fast neutron collisions, and the rates can be still further increased by using materials containing the reacting isotope in enriched proportions.

Values of the doses accurate to between 1% and 10% are readily obtained from the known cross sections and the integrated thermal neutron flux which can be obtained from the activity of monitoring cobalt foils placed with the specimens. The size of the samples is restricted by the necessity of removing heat liberated in the reactions and also, in the case of boron compounds especially,

by the strong absorption of thermal neutrons in the samples. The destruction of the reacting nuclei, and the accumulation of reaction products give impurity effects, which become large as the integrated thermal neutron fluxes approach the reciprocal of the reaction cross sections.

4.2. *Cyclotron Irradiation*

Mono-energetic beams of protons, deuterons and α-particles, with energy between 5 and 50 Mev and current density of 10^{-6} A cm^{-2} can be extracted fairly readily from small cyclotrons. This current density corresponds to a proton flux of $6{\cdot}24 \times 10^{12}$ particles cm^{-2} sec^{-1}. Convenient irradiation times are shorter than those of reactor irradiations, but the rate at which damage is introduced is usually considerably higher.

The techniques of cyclotron irradiation have been described in detail by Yockey *et al.* (1954). Doses may be measured to within about 10% by integrating the current flowing to the specimens by the Faraday cup method. To ensure that the bombarding particles are effectively mono-energetic, thin specimens must be used, which are not suitable for the measurement of some physical properties. The energy absorbed by the specimen renders close temperature control difficult. Nuclear transmutations are usually quite negligible.

Cyclotron produced neutrons have been used (Sidhu and Henry 1950, Hamilton, Putnam and Ehrmann 1952), but the fluxes available appear to be much lower than those obtainable in reactors.

4.3. *Van de Graaff Irradiation*

Van de Graaff machines have been used for electron irradiations. Beam currents of about 6×10^{14} electrons cm^{-2} sec^{-1} are readily available. The cross sections for primary knock-ons are about a thousand times less than those for cyclotron produced particles and Van de Graaff machines generally induce damage some ten to a hundred times slower than cyclotrons. The techniques are very similar to cyclotron irradiation, but since the electron ranges are greater than for heavy charged particles with the same energy, higher beam currents and thicker specimens can be used before insuperable heat transfer problems are encountered. γ-radiation is always present; transmutations are negligible.

The energies of the electrons are in the range 0·5 to 5 Mev, and the minimum electron energy required to produce a primary knock-on lies in this range for the majority of elements. The determinations of displacement energy E_d referred to in § 2.1.1 have all been made by measuring the damage as a function of electron energy, and extrapolating to zero damage. Results obtained by Klontz (1952) are shown at figure 4.1.

4.4. *Gamma Irradiation*

A 100 curie source provides γ-ray fluxes of about 10^{11} quanta cm^{-2} sec^{-1}. The cross sections for primary displacements are low and only very small fractions of displaced atoms can be produced in a convenient time. Dugdale (1955) has used γ-radiation from ^{60}Co, which has a mean energy of 1·3 Mev.

4.5. *Natural Radioactivity*

The spontaneous radioactivity of heavy elements produces α-particles of various energies in the range 4–10 Mev (Briggs 1954) which have been used to

produce irradiation damage. The fluxes obtainable are largest with the shortest lived elements when they can be concentrated in sufficient bulk. Probably the best available is polonium (half-life 140 days), which was used by Brattain and Pearson (1950). In principle, a flux of about 10^{11} α-particles cm^{-2} sec^{-1} is obtainable at the surface of a sufficiently thick source. Brattain and Pearson worked with a flux of 2×10^6 α-particles cm^{-2} sec^{-1}. Marsden (1910), using radon, worked with up to 10^9 α-particles cm^{-2} sec^{-1}. The α-particles only penetrate distances of about 10^{-3} cm, which places severe restrictions (with external sources) on the type of experiment that is feasible.

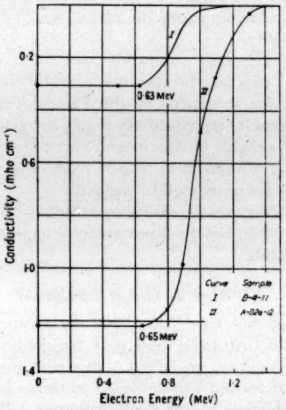


Figure 4.1. Extrapolation curves used in the determination of threshold energy in germanium. The electron irradiations were carried out at 77°K (Klontz 1952).

Very large doses have been produced in large volumes by the natural radioactivity of uranium and thorium contained in minerals, acting for times of between 10^8 and 10^9 years. In 10^9 years 15% of the ^{238}U originally contained in a mineral has decayed to lead in eight separate α-disintegrations, each of which produces about 2000 displaced atoms mainly due to the recoiling heavy nucleus. Consequently the atoms of a mineral 10^9 years old and containing 10% (atomic) of uranium have all been displaced on the average about 250 times. Doses in thorium-containing minerals are somewhat less because the half-life of thorium is longer, and only six α-disintegrations occur in its decay scheme. The range of the recoiling nuclei is very short ($\sim 10^{-6}$ cm, Bohr 1948), and so damage caused by radioactive inclusions is essentially due to the α-particles.

Minerals containing radioactive elements only as impurities exhibit a wide range of doses from one specimen to another, and it is possible to study effects as a function of dose (e.g. Hurley and Fairbairn 1953). Doses in radioactive minerals are measured by essentially the same methods as are used for age determinations, the most usual of which is the chemical estimation of Pb and U. Hurley and Fairbairn used a variety of methods to overcome the difficulties

presented by heterogeneous distribution of radioactivity, and the simultaneous presence of thorium. Their accuracy appears to be within 10% to 20%.

We summarize the main features of the various irradiation techniques with some illustrative figures in table 4.1.

Table 4.1. Summary of Irradiation Techniques

Radiation	Convenient energy (MeV)	Convenient flux (particles cm^{-2} sec^{-1})	Convenient time	No. of displacements per atom
Neutrons	2	10^{13}	1 month	10^{-1} in Cu
Slow neutron reaction ^{10}B (n, α)	^{7}Li, 0·87 ^{4}He, 1·53	10^{13}(neutrons)	1 month	2 in BN.
Deuterons	10	6×10^{12}	10 hours	10^{-2} in Cu
Electrons	2	6×10^{14}	10 hours	10^{-3} in Cu
γ-rays	1·3	10^{11}	1 month	10^{-8} in Cu
Natural sources (External)	5·3(α)	10^{3}	1 month	5×10^{-4} in Cu†
Natural sources (Minerals)	~6(α)	—	10^{9} years	100 in $ThSiO_4$

† Average over range of 10^{-3} cm

§ 5. Experimental Results

5.1. *Metals*

Except for changes in yield stress, the physical properties of metals are not greatly affected by irradiation at room temperature, largely because of the ease with which lattice defects anneal out. In aluminium, for instance, room temperature irradiation causes no observable changes of physical properties, including yield stress, although irradiation (and measurement) at liquid nitrogen temperatures gives marked alterations. Thus a study of irradiation phenomena in metals involves irradiation at the lowest possible temperatures in order to freeze in all the lattice defects. A large proportion of the measurements have been made on electrical resistivity which is, perhaps, the easiest physical property to measure under these conditions.

5.1.1. *Copper.*

Martin *et al.* (1951) bombarded copper at −150°C with 35 MeV α-particles and found an increase in electrical resistivity, the greater part of which annealed out at room temperature. Marx *et al.* (1952) using 12 MeV deuterons at −150°C measured resistivity changes as a function of integrated deuteron flux up to about 10^{17} deuterons cm^{-2}. They carried out annealing experiments which showed the recovery to have an activation energy of 0·20±0·05 ev at −110°C and about 1 ev at room temperature. More accurate measurements by Overhauser (1953) using the same irradiation conditions, showed that the activation energy rose from 0·2 ev as annealing proceeded and reached a steady value of 0·68 ev at −30°C, as shown in figure 5.1. When this recovery was complete no other well-defined activation energy was found at annealing temperatures up to 160°C. Eggleston (1953) using 35 MeV α-particles, found an activation energy of 0·72 ev between −65 and −20°C and a second well-defined activation energy of 2·12 ev between 250 and 300°C. In a further investigation of the lowest activation energy for recovery, Cooper *et al.* (1954) carried out deuteron irradiations at 12 and 16°K. On heating the copper after irradiation they observed a sharp drop of 50% in the excess resistivity at 40°K indicative of a unique activation energy of about 0·1 ev.

In all of these irradiations the specimens were sufficiently thin to allow complete penetration by the charged particles, and the damage consisted of small groups of interstitials and vacancies.

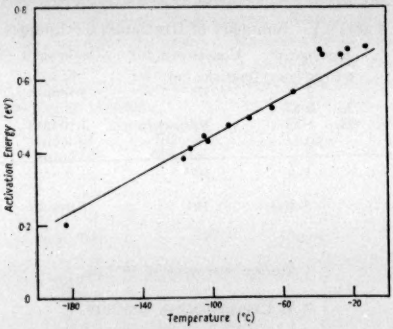


Figure 5.1. Activation energy for recovery in copper. The line extrapolates to zero activation energy at 0°K (Overhauser 1953).

Marx *et al.* (1952) suggested several possible reactions between interstitials and vacancies which could explain their early observations, and Overhauser (1953) ascribed the 0·7 ev reaction to the migration of vacancies. Huntington and Seitz (1942) and Huntington (1942) had previously calculated the migration activation energies for interstitials and vacancies as $\frac{1}{2}$ ev and 1 ev respectively; the more recent calculations of Huntington (1953) give $\frac{1}{4}$ ev for interstitials. Brinkman *et al.* (1954) have discussed in detail the migration of interstitials and vacancies. They associated the activation energy of 1·25 ev, found by Bowen, Eggleston and Kropschot (1952) for the annealing of cold-worked copper, with the migration of vacancies. (This association is supported by the activation energy of 1·2 ev for vacancy migration which can be deduced from electrical measurements of Meechan and Eggleston (1954) and the self-diffusion activation energy.) The lower activation energies found in irradiated copper were therefore associated with interstitial migration.

Because of distortion of the lattice in the neighbourhood of interstitials and vacancies, the activation energy for the annealing of close interstitial vacancy pairs will be appreciably less than that required for the migration of interstitials in the undistorted lattice (Marx *et al.* 1952, Overhauser 1953). Thus it is possible to argue that the marked recovery at 40°K is due to the annealing of interstitials in the positions closest to vacancies and that the activation energy rises as the average interstitial–vacancy separation increases, until 0·7 ev represents the activation energy for migration of interstitials in a lattice with small *average* strain. The kinetics of the 0·7 ev reaction observed by Overhauser and Eggleston differ from those of the expected second-order reaction, and the calculations of Overhauser show that the discrepancy can reasonably be attributed to lattice strain.

The measurements of Overhauser (1954) on the stored energy liberated by deuteron irradiated copper (0·052 cal g^{-1} for 10^{17} deuterons cm^{-2}) provide a partial

confirmation for the above interpretation. Between −150°c and room temperature he found that the ratio of stored energy released to resistivity recovered was constant, indicating that a single process (i.e. the return of interstitials to vacancies) was operative. In addition, McDonell and Kierstead (1954) have observed a decrease in the density of copper irradiated at −190°c by 21 mev deuterons. As we shall see, their results (shown in figure 5.2) suggest that interstitials are present at −190°c and that the majority are removed by annealing at room temperature.†

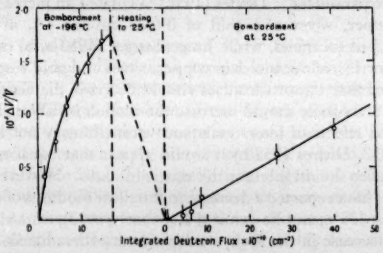


Figure 5.2. Specific volume increase induced by deuteron bombardment (McDonell and Kierstead 1954).

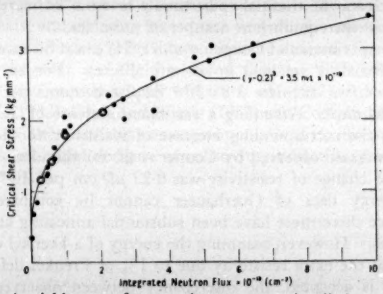


Figure 5.3. Change in yield stress of copper due to neutron irradiation at room temperature (Blewitt, private communication).

The increase in yield stress observed by Blewitt and Coltman (1951) (see figure 5.3) and McReynolds *et al.* (1954), and the increase in hardness observed by Billington and Siegel (1950) and Dixon and Meechan (1953) may be expected for almost any type of lattice defect. Redman *et al.* (1953) found an activation energy for recovery of 2·2 ev which is close both to the 2·12 ev observed by Eggleston (1953) in the resistivity recovery and to the activation energy of 2·02 ± 0·15 ev for self-diffusion (Kubaschewski 1950). Seitz (1952 b) suggested that interstitials and vacancies may coagulate into groups which would be removed by a self-diffusion process, and it is possible that such groups may be responsible

† *Note added in proof.* More recent measurements (Kierstead 1954) suggest that the volume changes may be five times smaller than the changes reported by McDonell and Kierstead (1954).

for the observed phenomena. There is, however, no positive evidence for the nature of the defects which recover with 2·1 ev activation energy.

In addition to the theoretical calculations, already mentioned, of the displacement energy and migration activation energies in copper, attempts have been made to predict the effect of lattice defects on the physical properties. Tucker and Sampson (1954) conclude that 1% of interstitials will give rise to a 1% increase of lattice parameter (3% decrease in density) while 1% of vacancies will give $\frac{1}{5}$% decrease of lattice parameter. Dexter (1952) calculated an increase in the electrical resistivity of copper, silver and gold of 0·6$\mu\Omega$ cm for 1% of interstitials and 0·4$\mu\Omega$ cm for 1% of vacancies, while Jongenburger (1953 a, b) calculated 1·3, 1·5 and 1·5 $\mu\Omega$ cm for 1% of vacancies in copper, silver and gold respectively. Dienes (1952 a) estimated that 1% of vacancies should decrease the elastic moduli by 1% and that 1% of interstitials should increase the moduli by about 10%. While the magnitudes of the effects of interstitials and vacancies may not be so different as this (Nabarro 1952, Dienes 1952 b) it would appear that equal numbers of interstitials and vacancies should increase the elastic moduli. However, Dieckamp and Crittenden (1954) have reported a decrease in the shear modulus of copper irradiated and measured at −175°c, and Bowen and Rodeback (1953), from electrical measurements, found a decrease in the Debye temperature of irradiated copper.

Witzig (1952) observed no change in the creep rate of copper irradiated with 16 Mev deuterons at 260°c; his calculations showed that little change would be expected either from the thermal spike model or from an increase, based on the collision model, of the equilibrium number of vacancies.

Only in the experiments of Cooper *et al.* (1954) can it be assumed that most of the displaced atoms are retained in interstitial sites. For an average deuteron energy of 10 Mev, we calculate $3{\cdot}8\times10^{-3}$ displaced atoms per atom for a dose of 10^{17} deuterons cm^{-2}. Assuming a resistance increase of 2$\mu\Omega$ cm for 1% of Frenkel defects, the corresponding increase of resistivity should be 0·8 $\mu\Omega$ cm. The resistivity increase observed by Cooper *et al.* for this dose was 0·19 $\mu\Omega$ cm; the initial rate of change of resistivity was 0·23 $\mu\Omega$ cm per 10^{17} deuterons cm^{-2}. The stored energy data of Overhauser cannot be compared directly with calculations, since there must have been substantial annealing at the temperature of his irradiations. However, assuming the energy of a Frenkel defect to be 5 ev, he concludes that the extra resistivity due to 1% of Frenkel defects is 11$\mu\Omega$ cm. If this estimate is accepted, the discrepancy between observed and calculated resistivity changes in the experiments of Cooper *et al.* is increased to a factor of 16.

The fraction of displaced atoms in the volume of the specimen considered by McDonell and Kierstead should be about 3×10^{-3} per 10^{17} deuterons cm^{-2}, corresponding to a fractional volume increase of 9×10^{-3}. Taking into account the annealing observed by Cooper *et al.*, this agrees quite well with the observed volume increase (initially $2{\cdot}5\times10^{-3}$ per 10^{17} deuterons cm^{-2}).†

The picture which we have of the recovery is therefore as follows. The first stage is the annealing of close pairs with activation energies rising from 0·1 to 0·7 ev; this is followed by almost free diffusion of randomly distributed interstitials with an activation energy of 0·7 ev. During this second stage some groups of interstitials are formed and some interstitials escape to dislocations or grain boundaries, leaving a small excess of vacancies, the removal of which constitutes

† See footnote on previous page.

the third stage. Finally, the groups of interstitials are eliminated by a self-diffusion process, recovery being complete at 400°c. Qualitatively, this picture is in reasonable accord with the observations, but the quantitative agreement is not so satisfactory. We have ignored other possible processes, such as the migration of vacancy pairs (Bartlett and Dienes 1953) and the dissociation of groups of defects, which may be important. Our interpretation is not the only possible one (Broom 1954) and its validity can only be determined by further experimental work.

5.1.2. *Copper–gold.*

A number of irradiation experiments have been carried out with the copper–gold alloy Cu_3Au. This is a typical order–disorder alloy of which a description is given in the review article by Nix and Shockley (1938). For this alloy the Bragg–Williams definition of the degree of order S is $S=4(r-\frac{1}{4})/3$ where r is the fraction of gold atoms occupying ' gold sites ' i.e. lattice sites which are occupied by gold atoms when the alloy is perfectly ordered. $S=1$ for perfect order and $S=0$ for disorder. The equilibrium degree of order is zero above the critical temperature (388°c) and increases to unity as the temperature is lowered to 0°K. At room temperature the times necessary to reach equilibrium are so long that all the ordered alloys have appreciably less than the equilibrium of order. If the alloy is quenched after prolonged heating just below the critical temperature, domains of long-range order are present and ordering can proceed readily at 150°c, but if quenched from above the critical temperature, no such domains exist and ordering takes place very much more slowly.

The experiments of Siegel (1949) showed that the ordered alloy was disordered by fast neutron irradiation, and the disordered alloy was little affected. Similar experiments have been carried out by Martin *et al.* (1951), Adam and Dugdale (1951), Glick *et al.* (1952), Glick and Witzig (1953) and Cook and Cushing (1953). The main feature of these subsequent experiments is that, except in high fluxes of fission neutrons, an initial decrease of resistivity, indicative of ordering, can be observed (see figure 5.4).

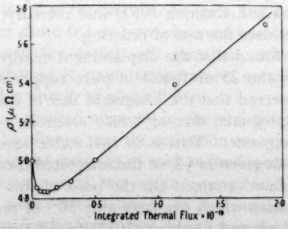


Figure 5.4. Change in resistivity of Cu_3Au due to neutron irradiation (reactor spectrum) at room temperature (Dugdale, private communication).

The experiments of Blewitt and Coltman (1952, 1954) and of Dixon and Bowen (1954) show that the rate of ordering is considerably enhanced *during* irradiation by neutrons and protons, and it is suggested by Blewitt and Coltman that the ordering is due to the diffusion of lattice defects. This suggestion is confirmed by the increased rates of ordering observed *after* short irradiations (producing

little or no disorder) by electrons (Adam, Green and Dugdale 1952, Dixon, Meechan and Brinkman 1953), γ-rays (Dugdale 1955) and neutrons (Coltman and Blewitt 1953). Brinkman *et al.* (1954) argued that it is improbable that the alloy can be ordered by the movement of interstitials, and they postulated that the ordering is due to vacancy migration. Their qualitative arguments are supported by the measurements of Dugdale and Green (1954) who found that the activation energies for ordering were almost the same (0·9 ev) for cold worked, quenched and electron irradiated specimens, since an excess of vacancies would be expected in the first two cases. Brinkman *et al.* (1954) irradiated ordered and disordered specimens with protons at a temperature below -100°c and observed resistivity increases of 4·25 and 0·65 $\mu\Omega$ cm respectively. The specimens were then heated, and between -60 and 0°c both recovered by 0·60 $\mu\Omega$ cm. At higher temperatures the resistivity of the ordered specimen decreased by a further 0·45 $\mu\Omega$ cm and that of the disordered specimen decreased by 0·05 $\mu\Omega$ cm. They attributed the low temperature recovery to migration of the interstitials, some of which diffuse to dislocations and grain boundaries, and the subsequent recovery of the ordered specimen to migration of the surplus vacancies. The lack of nuclei in the disordered specimen explained the absence of ordering.

The initial decrease of resistivity which occurs during room-temperature neutron irradiation of ordered specimens can be explained in terms of vacancy migration. As discussed in § 2.3, each fast neutron collision will give rise to a highly disordered region. After the migration of interstitials, the surplus vacancies assist further ordering of the surrounding matrix which, although termed ordered, has not the equilibrium degree of order at temperatures below about 200°c. As the number of disordered regions increases, the migration distance of vacancies decreases (Blewitt and Coltman 1954) and this, together with the increasing order of the matrix, leads to a saturation of the order, after which disordering predominates. In high fluxes, the vacancies have no time to move before their ordering efficiency is seriously reduced. Because practically every type of radiation which would be expected to produce displaced atoms has been observed to enhance the rate of ordering of Cu_3Au, the above interpretation seems preferable to the assumption (Cook and Cushing 1953) that mercury, produced by thermal neutron absorption, increases the rate of ordering.

Dugdale (1955) has found that the displacement energy for copper in Cu_3Au is appreciably less than the 25 ev found in pure copper and is probably about 10 ev. He has also observed that the integrated flux of cobalt γ-rays is approximately 200 times the integrated flux of 1 MEV electrons needed to produce the same increase in ordering rate. This is in reasonable agreement with the ratio, calculated by the methods given in §2, of the integrated fluxes needed to produce a given fraction of displaced atoms. On the basis of this data, ordering can be observed after an irradiation which displaces only 10^{-9} of the atoms.

Siegel (1949) and Cook and Cushing (1953) found that regions of about 10^4 atoms are disordered at each fast neutron collision and Seitz (1949) has calculated (using $E_d=25$ ev) that the number of displaced atoms would be an order of magnitude less. He was therefore led to postulate the existence of high temperature thermal spikes to explain the observed disordering. For this alloy, in which the atomic weights are very different, calculations are difficult, but it would appear that the replacement collisions used (Kinchin and Pease 1955) to explain the disordering of a similar alloy (Ni_3Mn) offer an alternative explanation.

Fillnow *et al.* (1953) have found an increase in hardness in both ordered and disordered Cu_3Au, analogous to the hardness changes observed in copper.

The model which we used to explain irradiation effects in copper seems to be applicable to the case of Cu_3Au, taking into account the smaller values of displacement and activation energies.

5.1.3. *Other metals and alloys.*

(*a*) *Electrical properties.* Marx *et al.* (1952) measured resistivity changes in silver, gold, nickel and tantalum irradiated with deuterons at −150°C. The largest effects were observed in tantalum, and the rate of increase of resistivity fell off with irradiation dose in all cases. Larger changes were observed in silver and gold at 12°K by Cooper *et al.* (1954). Their results for copper, silver and gold are summarized in table 5.1, in which changes observed for an integrated flux of 10^{17} deuterons cm^{-2} are compared with those calculated from the initial rate of change of resistivity; even at 12°K there appears to be some annealing.

Table 5.1. Change in Resistivity ($\mu\Omega$ cm) for 10^{17} deuterons cm^{-2}. Deuteron energy 12 MeV

		Copper	Silver	Gold
12°K	Extrapolated from initial slope	0·23	0·26	0·38
	Observed	0·19	0·20	0·31
135°K	Extrapolated from initial slope	0·09	0·14	0·26
	Observed	0·06	0·10	0·19

On raising the temperature after irradiation, the resistivity of silver dropped sharply at about 30°K by 20% of the radiation-induced increase (compared with the 50% drop at 40°K in copper), while for gold no sharp drop was observed. Assuming in all cases 25 ev for the displacement energy and $2\mu\Omega$ cm for the resistivity due to 1% of Frenkel defects, the calculated increases of resistivity for copper, silver and gold are 0·8, 1·1 and 1·6 $\mu\Omega$ cm respectively. The observed resistivity changes show the same trend as the calculated values, but are appreciably smaller.

Activation energies of about 0·2 ev were observed by Marx *et al.* for silver and gold, as for copper; even for nickel and tantalum, which retained a much larger fraction of the change at room temperature, annealing was observed at −110°C with an activation energy of about 0·3 ev.

Dugdale (1952) has measured changes in the resistivity of platinum due to neutron irradiation and has shown that the activation energy for recovery (1·19 ev) is close to that observed in the recovery of cold worked specimens. The measurements were made only over a small temperature range and it was not shown that the activation energies were ' well-defined ' in both cases. There is, however, a strong presumption that both effects were due to vacancy migration.

Randolph (1952) found an increase in the resistivity of molybdenum using 10 MeV deuterons, and Kinchin (see Dugdale 1955) has measured the increase of resistivity of molybdenum and tungsten bombarded with fission neutrons at a temperature of about 30°C (figure 5.5). The activation energy for recovery in molybdenum rises from 1·20 to 1·35 ev during annealing at temperatures up to 190°C. The apparent order of reaction (§ 3) then changes and the activation energy reaches a steady value of 1·47 ev between 230 and 300°C, at which

temperature 90% of the change has recovered. Only 20% of the change in tungsten recovers at temperatures up to 300°c and the activation energy rises from 1·2to over 1·6 ev. Resistivity changes in aluminium irradiated by neutrons at low temperatures (McReynolds *et al.* 1954) anneal out completely between −80 and −20°c.

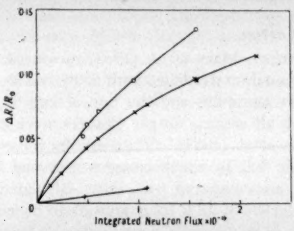


Figure 5.5. Fractional resistance increase of metals irradiated with fission neutrons at 30°c. o tungsten; × molybdenum; + platinum (Kinchin).

Wruck and Wert (1954) irradiated iron, cobalt and nickel with 10 Mev deuterons at −150°c and found that the increase of resistivity in iron is about five times greater than that in cobalt or nickel. Neutron irradiation also gives a larger change in iron and the authors attribute this to the difference in structure.

Aronin (1954) has measured changes of resistivity and of magnetic properties in the order–disorder alloy Ni_3Mn after neutron irradiation. From his results he calculated that each fast neutron collision gave rise to about 5000 random replacements. The picture, used by Cook and Cushing (1953), of islands of disorder in an ordered matrix is more applicable to the case of fast neutron bombardment, and this picture requires about 10^4 random replacements per fast neutron collision. The number calculated by Kinchin and Pease (1955) was 7450, of which 1350 were due to the random return of displaced atoms and 6100 were due to replacement collisions. Eggleston and Bowman (1953) have observed large increases in the resistivity of β-brass irradiated with α-particles at −100°c, with complete recovery below room temperature; the effects were attributed to disordering. Rosenblatt *et al.* (1954) found that the resistivity of α-brass *decreased* during neutron irradiation and that the resistivity recovered its original value after annealing at 170°c. This behaviour is thought to be due to an ordering of the alloy.

Changes in the thermoelectric power of iron and constantan have been reported by Andrew and Davidson (1953).

(*b*) *Density.* Tucker and Sampson (1954) have reported lattice expansions of the order of 5×10^{-2}% in various neutron irradiated metals and Adam (see Dugdale 1955) has found an increase of $1{\cdot}5 \times 10^{-2}$% in the lattice spacing of molybdenum after an irradiation of $2{\cdot}3 \times 10^{19}$ fission neutrons cm^{-2} at 30°c. These small changes suggest that, because of rapid recovery, relatively few interstitials remain after room temperature irradiation.

(*c*) *Mechanical properties.* Increases in yield stress or hardness have been found in aluminium at low temperatures (McReynolds *et al.* 1954), molybdenum (Geib and Grace 1952), iron, zinc (Kunz and Holden 1954), mild steel (Meyer 1954) and stainless steels (Billington and Siegel 1950). In iron, as in copper,

the activation energy for recovery (3·1 ev) is close to the activation energy for self-diffusion.

(*d*) *Diffusion and creep.* Johnson and Martin (1952) looked for a change in the self-diffusion coefficient of silver under proton bombardment, and found no appreciable change at the lowest temperature of measurement (555°c). Their calculations showed that the equilibrium concentration of vacancies produced by the radiation would be expected to be small compared with the vacancy concentration in thermal equilibrium.

These results may not appear at first sight to be consistent with the explanation of, for instance, the enhanced ordering rates in Cu_3Au, where the concentration of vacancies due to radiation is supposed to exceed the equilibrium concentration. This is because the diffusion lengths needed to measure bulk diffusion coefficients experimentally are so great that this type of experiment cannot be carried out in a temperature range where radiation effects produce a significant increase in the equilibrium number of vacancies or interstitials.

The experiments of Andrade (1945) on the effect of α-particle bombardment on the creep rate of cadmium have been repeated by Makin (1955) who found no significant change in creep rate. Jones, Munro and Hancock (1954) found no change in the creep rate of aluminium due to neutron irradiation.

(*e*) *Phase transformations and precipitation.* Fleeman (1953, 1954) has found that the transformation from white to grey tin is nucleated by neutron irradiation, and Denney (1953, 1954 a) showed that iron precipitated in copper is transformed from the face-centred cubic to the stable ferromagnetic body-centred cubic structure under the influence of electron irradiation. The threshold found by Denney (0·5 Mev) gives a lower limit of 25 ev to the displacement energy for face-centred cubic iron. The transformation could be induced by vacancies formed in the copper matrix, for which the threshold energy is known to be 0·49 Mev, or by the displacement of iron atoms. Denney (1954 b) subsequently found that proton irradiation causes the ferromagnetism of the transformed precipitate to *decrease* and interpreted this experiment in terms of thermal spikes. Solution of the precipitate by replacement collisions may, however, be the explanation.

Billington and Siegel (1950) studied the effects of fast neutron irradiation on supersaturated solutions of beryllium in copper. More extensive measurements by Murray and Taylor (1954) showed that the changes in resistivity and hardness are very similar to those produced by ageing at temperatures of about 100°c and they are thought to be due to the formation of precipitate nuclei. The similarity was checked by retrogression experiments (solution of the nuclei) at higher temperatures. Latent changes, which become apparent on subsequent heating, are introduced by irradiation at low temperatures (Cleland, Billington and Crawford 1953). These experiments suggest that the precipitation is nucleated by the migration of lattice defects (probably vacancies). Density changes observed by Murray and Taylor were not explained.

5.2. *Semiconductors*

5.2.1. *Germanium.*

Since a typical sample of n-type germanium may contain only 10^{-7} free electrons per atom and each displaced atom removes of the order of one free electron, it is evident that the electrical properties of semiconductors are extremely sensitive to radiation.

The changes of conductivity of n-type (electron excess) and p-type (electron deficit or hole excess) germanium due to deuteron irradiation are shown in figure 5.6 (Fan and Lark-Horovitz 1955). It will be seen that the conductivity alters by several orders of magnitude, in contrast with the relatively small changes in metals. Changes of conductivity and Hall coefficient have been extensively

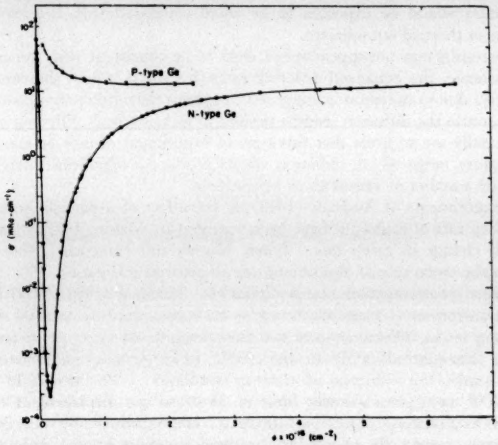


Figure 5.6. Change in conductivity of deuteron-irradiated germanium as a function of dose (Fan and Lark-Horovitz 1955).

studied using irradiation with heavy charged particles (Lark-Horovitz, Bleuler, Davis and Tendam 1948, Brattain and Pearson 1950, Forster, Fan and Lark-Horovitz 1952) and neutrons (Davis, Johnson, Lark-Horovitz and Siegel 1948, Johnson and Lark-Horovitz 1949, Crawford and Lark-Horovitz 1950, Lark-Horovitz 1951, Cleland *et al.* 1951 a, b). The effects of nuclear transmutations during neutron irradiation have been shown to be small, in agreement with calculations based on the known cross sections (Cleland, Lark-Horovitz and Pigg 1950). Measurements of Hall coefficient demonstrate that n-type germanium is converted to p-type by irradiation, and confirm that the irradiation effects are due largely to change in the numbers of charge carriers and not to any changes in mean free path. Stevens, Cleland and Crawford (1954) have measured changes in diamagnetic susceptibility corresponding to the changes in electron concentration.

James and Lark-Horovitz (1951) have given an explanation of the effects of neutron irradiation. They consider that vacancies act as electron traps (sometimes called acceptors) and conclude that each vacancy produces two localized empty levels in the forbidden band as shown in figure 5.7. Similarly interstitials act as donors of electrons (sometimes called hole traps) and each interstitial produces two localized filled levels, also shown in figure 5.7. Each interstitial–vacancy pair therefore produces four localized levels and two electrons. If n-type germanium, with an excess of electrons in the conduction band, is irradiated, the lower empty

levels become filled, and each interstitial–vacancy pair removes two electrons. If the material initially has a large excess of electrons, the conductivity is almost proportional to the number of electrons and the conductivity therefore decreases linearly. As the conduction band is emptied, with a corresponding depression of the Fermi level, the higher donor level ceases to be fully occupied, and at the point of minimum conductivity, when the Fermi level is near the middle of the forbidden band, only one electron is trapped per interstitial–vacancy pair. This agrees with the observation that the minimum conductivity occurs at approximately twice the dose which would be expected from a linear extrapolation of the early part of the conductivity–dose curve.

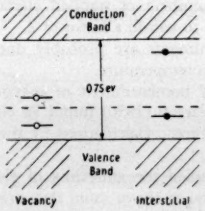


Figure 5.7. Energy levels in irradiated germanium (*after* James and Lark-Horovitz 1951).

On irradiation of high conductivity p-type germanium, with a large excess of positive holes in the conduction band, electrons from the interstitials drop into the conduction band and reduce the number of charge carriers, initially at the rate of two positive holes per interstitial–vacancy pair (see figure 5.6).

After long irradiations, when the number of defects greatly exceeds the initial electron excess or deficit, it would be expected that the Fermi level would reach a limiting position, independent of the nature of the starting material, approximately midway between the second and third highest levels above the valence band shown in figure 5.7. Experiments with neutron irradiation (Cleland *et al.* 1951 b) have shown that the limiting position of the Fermi level is between 0·1 and 0·17 ev above the top of the valence band.

This explanation can be only a qualitative account of the behaviour of germanium irradiated by neutrons or heavy charged particles, since more recent experiments (Forster, Fan and Lark-Horovitz 1953, Crawford *et al.* 1953) have shown that annealing after irradiation at temperatures below $-140°$c is accompanied by a redistribution of the localized energy levels, which has been tentatively ascribed to the formation of clusters of defects. The energy levels observed after room temperature irradiation cannot therefore be associated with single interstitials and vacancies.

In the case of irradiation by electrons of moderate energy, when isolated interstitial–vacancy pairs are produced, Fan and Lark-Horovitz (1955) have shown that there is no alteration of the energy levels on annealing and only the number of trapping centres is changed. They have found three localized energy levels in electron irradiated germanium.

The displacement energy for germanium (31 ev) has been determined by Klontz and Lark-Horovitz (1951, 1952). Klontz (1952) also found $dn_e/d\phi$, the number of electrons trapped per cm^3 per unit integrated electron flux. His

results, together with those of Brown, Fletcher and Wright (1953) and Cleland *et al.* (1951 a) are compared in table 5.2 with the numbers of displaced atoms per cm³ per unit integrated flux $dN/d\phi$ calculated by the methods given in § 2.

Table 5.2. Observed and Calculated Damage Rates in Germanium

Author	Radiation	Temperature	$dn_e/d\phi$	$dN/d\phi$
Klontz	1·5 Mev electrons	77°K	1·1	1·2
Klontz	1·5 Mev electrons	Room	0·6	1·2
Brown *et al.*	3 Mev electrons	Room	1·4	2·6
Cleland *et al.*	Pile neutrons	Room	3·2	7

Assuming that two electrons are trapped by each interstitial–vacancy pair, the observed and calculated numbers of trapped electrons in the low-temperature experiment of Klontz agree to within a factor of 2. The greater discrepancies for room temperature bombardments are probably due to the annealing which is known to occur below room temperature.

In the initial stages of bombardment of n-type germanium with 5·3 Mev α-particles, Brattain and Pearson (1950) found 78 electrons trapped per incident α-particle at room temperature. Calculations by the methods of § 2 give 85 displaced atoms per α-particle.

Brown *et al.* (1953) studied the annealing of electron irradiated germanium and found two annealing stages above room temperature, which they interpreted in terms of the theories (§ 3) developed by Fletcher and Brown (1953), assuming vacancies to be more mobile than interstitials. The measurements of Mayburg (1954) seem to show that interstitials are the more mobile. Brown *et al.* concluded that the first stage, with an activation energy of 1·58 ev, was due to the monomolecular recombination of very close pairs, and that the second stage, with an activation energy of 1·78 ev, was a ' liberation ' process of the kind described in § 3.1.1.

Florida, Holt and Stephen (1954) have found that neutron irradiation causes a decrease in the hole life-time in germanium and suggest that the lattice defects are recombination centres.

5.2.2. *Silicon.*

The effects of irradiation on silicon are in general similar to those on germanium. The limiting position of the Fermi level appears to be close to the middle of the forbidden band, so that both n-type and p-type silicon finish, after long neutron irradiation, with very high resistivity. The trapping levels appear to lie at about 0·7 to 0·8 ev below the conduction band, and this is confirmed by the observation of a strong infra-red absorption band at 0·7 ev (Johnson and Lark-Horovitz 1949, Lark-Horovitz, Becker, Davis and Fan 1950, Forster, Fan and Lark-Horovitz 1952).

Measurements of the specific heat of neutron-irradiated silicon by Keesom, Lark-Horovitz and Pearlman (1952) show that the linear electronic contribution vanishes in accordance with the electrical measurements, while the Debye temperature decreases. This suggests a decrease of elastic moduli which measurements by Binnie and Liebschutz (1954) seem to confirm.

5.2.3. *Indium antimonide.*

Indium antimonide has a narrow ($\sim$ 0·2 ev) energy gap between valence

and conduction bands and a high electron mobility. Irradiation by neutrons (Cleland and Crawford 1954 a, b, c) and electrons (Pepper *et al.* 1954) reduces the mobility and changes the number of charge carriers. The distribution of localized energy levels is such that the end product of long irradiation is high resistivity p-type material. The interpretation of the results is more complex than for germanium, since indium antimonide contains two types of atom.

5.2.4. *Graphite.*

Graphite probably has a slightly overlapping band structure (Kinchin 1953) and is intermediate in properties between metals and semiconductors. The changes in Hall coefficient due to neutron irradiation at 30°C are shown in figure 5·8 and can be explained in terms of trapping of electrons by lattice defects. In

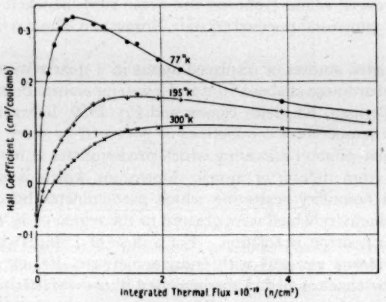


Figure 5.8. Hall coefficient of graphite irradiated with neutrons (reactor spectrum) at 30°C. • measured at 77°K; + measured at 195°K; × measured at 300°K (Kinchin 1954).

this case the changes in mean free path are appreciable, and the combination of decreasing mean free path and increasing carrier (positive hole) concentration leads to a saturation of the resistivity change (Kinchin 1954). The Hall coefficient changes lead to an estimate of 1 trapped electron cm^{-3} per neutron cm^{-2}. The experimental results suggest that the localized energy levels are not fully occupied, and there is considerable uncertainty in estimating the number of displaced atoms from the number of trapped electrons. The calculated (§ 2) number of displaced atoms is also very rough, since it is virtually proportional to L_C, but is within an order of magnitude of the experimental estimate.

5.3. *Covalent Compounds*

5.3.1. *Quartz.*

Coloration, embrittlement and fracture of quartz and glasses due to bombardment from naturally radioactive sources were observed many times by early workers (Rutherford, Chadwick and Ellis 1930, Lind and Bardwell 1923 a). Many studies have since been made of the coloration effects (e.g. Choong 1945). It is necessary here, as with all insulators, to distinguish clearly between ionization effects and effects due to displaced atoms. Pile irradiation produces relatively rapid changes in the optical absorption coefficients (Mitchell and Paige 1954), paramagnetic

resonance absorption (Griffiths, Owen and Ward 1954, 1955) and elastic constants (Johnson and Pease 1954) which either cease or continue much more slowly after a short dose. Very similar effects are observed by irradiation with X-rays, and in this case the coloration shows a well-marked saturation (Forman 1949). Johnson and Pease find the ionization doses to half-saturation to be of the order of 0·05 ev per molecule in both cases. It is a characteristic of these ionization effects that they vary widely from one specimen to another, and often from one part to another of the same specimen, and they are attributed to the trapping of electrons at defects already present in the crystal. Direct experiment (Brown and Thomas 1952) strongly suggests that, as supposed by early workers, these defects are impurities; this view is disputed by Dainton and Rowbottom (1954) who suggest that they may be dislocations or vacancies. The defects are hardly likely to be vacancies or interstitials, since fast neutron bombardment bleaches the ionization-induced colour (Johnson and Pease 1954, Mitchell and Paige 1954). In very carefully grown and purified crystals, Brown and Thomas find no ionization colouring.

The first detailed studies of displaced atoms in α-quartz were made by using fast neutron bombardment at about 50°C and low-temperature thermal conductivity measurements (Berman, Klemens, Simon and Fry 1950, Berman 1951, Klemens 1951). The low temperature conductivity is governed by three types of phonon scattering: phonon–phonon scattering which predominates at temperatures above 10°K, scattering from defects of atomic dimensions which is most effective at about 10°K, and boundary scattering which predominates below 10°K. Large decreases of conductivity which were greatest in the region of 10°K (see figure 5.9) were found upon neutron irradiation. For a dose of 1 unit (which we estimate was 10^{18} fast neutrons per cm^2 with energies between 100 ev and 2 Mev in a spectrum given by equation (4.1)) the increased phonon scattering was attributed to about 10^{-4} single defects per atom. The increases in phonon scattering observed at lower temperatures were attributed to increases of boundary scattering due to the presence of relatively large defects. On the implied assumption that these defects contain about 100 atoms, Klemens deduced that their concentration was about 10^{-7} per atom.

At the highest dose used, which corresponds to about 2×10^{19} neutrons cm^{-2}, the curve of thermal conductivity against temperature is much more similar to that of fused quartz than to that of α-quartz, and the concentration of defects is sufficiently high to produce a state which Klemens referred to as a ' dilute glass '.

The effects of heat treatment of the irradiated quartz were also studied, two main results being obtained. First the stability of the defects is considerable: heating at 100°C for three weeks produced no observable recovery, and heating well above the α–β transition point (573°C) was required before a major proportion of the conductivity change was recovered. Secondly, the large defects annealed more rapidly than the single interstitials and vacancies, although as Berman points out, the same effects might arise by the splitting up of the large defects during heat treatment.

X-ray examination of the effects of neutron bombardment on various forms of quartz provides a second main body of data. Wittels (1953) and Wittels and Sherrill (1954), whose results are confirmed qualitatively by Primak, Fuchs and Day (1953), show there is a transition to an optically isotropic glassy state at doses of about 10^{20} neutrons cm^{-2}. Beyond this dose, no single-crystal X-ray pattern

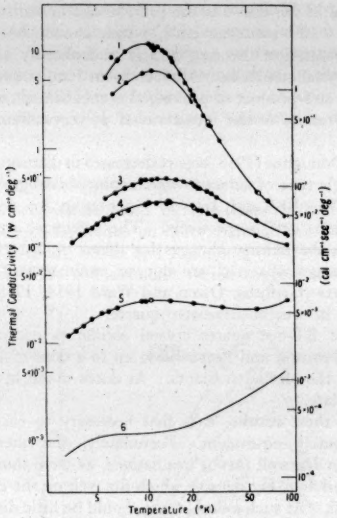


Figure 5.9. Thermal conductivity of neutron irradiated quartz crystal perpendicular to the axis. Cumulative irradiation dose: 1, unirradiated; 2, 0·03 units; 3, 1 unit; 4, 2·4 units; 5, 19 units. Curve 6 shows the thermal conductivity of quartz glass (Berman 1953).

was observed. At lower doses, where the x-ray reflections are still sufficiently sharp, Wittels observed an anisotropic expansion of the structure of α-quartz, the expansion along the *a*–axis being three times (Johnson and Pease 1954) greater than that along the *c*–axis. As Wittels points out, the largest interstitial sites available are irregular channels parallel to the *c*-axis, and the anisotropic expansion is qualitatively well explained by the supposition that atoms are displaced to such sites. Density changes of α-quartz as a function of dose are shown in figure 5.10.

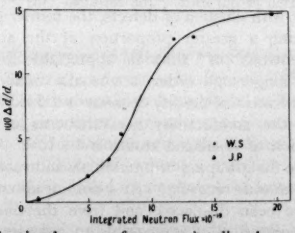


Figure 5.10. Fractional density decrease of neutron-irradiated α-quartz (room temperature). W.S., hydrostatic measurements of Wittels and Sherrill (1954); J.P., x-ray measurement of Johnson and Pease (1954).

A remarkable feature of the curve is the pronounced upward trend of the slope at doses of about 7×10^{19} neutrons cm^{-2}, which cannot be accounted for by any of the simple saturation theories, and is undoubtedly associated with the breakdown of the crystal structure. All four room temperature forms of quartz have the same glassy end-product of prolonged bombardment, and heat treatment (16 hours at 930°c) results in the formation of polycrystalline α-quartz in each case.

McClelland and Donoghue (1953) report decreases in diamagnetic susceptibility. Assuming only a single type of defect to which spin $\frac{1}{2}$ is assigned, they find defect concentrations of $0{\cdot}78\times10^{-3}$ and $1{\cdot}1\times10^{-3}$ per atom for doses of $0{\cdot}51\times10^{19}$ and $3{\cdot}8\times10^{19}$ neutrons cm^{-2} respectively. These changes are non-linear in the opposite direction to the density changes for doses smaller than 10^{20} neutrons cm^{-2}. That the changes observed are due to paramagnetic defects is shown directly by the reports (Griffiths, Owen and Ward 1954, 1955) of paramagnetic resonance absorption in neutron-irradiated quartz.

The frequency of BT-cut quartz crystal oscillators decreases linearly with neutron irradiation (Johnson and Pease 1954) up to a dose of 10^{19} neutron cm^{-2}, due to a decrease in the elastic constants. At doses much in excess of this the crystals cease to oscillate.

To intercompare these results, it is first necessary to ensure that the dose scales are approximately equivalent. Fortunately, the thermal conductivity studies were based on Harwell (BEPO) irradiations, as were those of Johnson and Pease; these latter find an x-ray density which fits well on the density–dose curve of Wittels and Sherrill. At such a dose, there should be little discrepancy between hydrostatic and x-ray density measurements. Consequently, these three studies are based on approximately the same dose scale which we use in the rest of this section. About 150 atoms will be displaced by each fast neutron collision on this dose scale. McClelland and Donoghue use doses of neutrons with energy greater than 0·5 MeV, and their doses may well be effectively five times greater. This would account for the discrepant non-linearity referred to above.

No detailed theory exists which provides a quantitative account of the nature of the irradiation damage, nor is there available an experimental value of E_d. However, the general picture is fairly clear. At very low doses (in the centre of BEPO, $\lesssim10^{16}$ neutrons cm^{-2}) ionization effects predominate. From this dose to about 2×10^{19} neutrons cm^{-2} there is a steady accumulation of damage, apparently roughly linear with dose, which probably consists of isolated interstitials and vacancies, together with clusters of defects, the former greatly predominating in number and affecting a greater proportion of the atoms. X-ray powder photographs at 10^{19} neutrons cm^{-2} show no appreciable line broadening, and it may be inferred that long-range order is virtually unaffected. At a dose of 10^{19} neutrons cm^{-2}, the fractional density decrease is $2{\cdot}3\times10^{-3}$, the proportion of isolated defects from the conductivity measurements is about 1×10^{-3}, and the calculated proportion of displaced atoms is 4×10^{-3}. This is quite a satisfactory agreement, since the grouping of interstitials and vacancies is likely to produce appreciable instantaneous recovery. At a dose of about 10^{22} neutrons cm^{-2}, 4% of the atoms have been displaced, and from the discussion in § 3.4 one would expect the concentration of interstitials to saturate at about this dose. Instead, however, the long range order breaks down (without producing any macroscopic shattering). From the non-linear upward trend of the density

curve, it appears as if the structure becomes unstable to further accumulation of interstitial atoms, and the damage in this region is a form of phase change to the metamict state. We discuss this change to the metamict state in a later section. The position has been further complicated in the case of quartz by Hurley and Fairbairn's (1953) report that quartz is not found metamict naturally, even when in the presence of radioactivity sufficient to turn other minerals metamict.

5.3.2. *Thermal, optical and magnetic properties.*

In spite of the success achieved by low-temperature thermal conductivity studies of irradiated quartz, results on other materials are limited to the recently reported results of Berman, Foster and Rosenberg (1955) with neutron-irradiated sapphire (see figure 5.11) and diamond which suggest similar effects to those

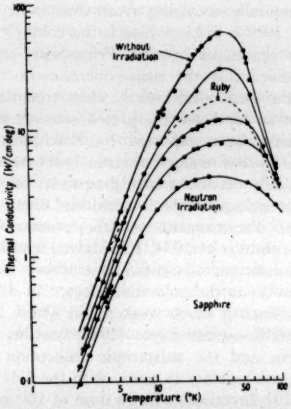


Figure 5.11. Thermal conductivity of sapphire as a function of temperature. Upper curve is for unirradiated sapphire. Lower curves for sapphire irradiated with reactor spectrum neutrons at room temperature for integrated thermal fluxes of 0·15, 0·9, 2 and 5×10^{18} neutrons cm^{-2}, from top to bottom (Berman, Foster and Rosenberg 1955).

found in quartz. Stored energies of 100 cal g^{-1} in SiC and diamond, and changes in the specific heat of quartz are referred to by Primak (1954).

Large numbers of observations have been made on coloration, luminescence and phosphorescence of irradiated minerals and glasses. In many cases the changes observed are predominantly ionization effects. We shall deal with coloration effects only in diamond, where a number of carefully controlled experiments have been reported, and where the role of ionization effects is certainly small in most cases.

(*a*) *Coloration in diamond.* It was shown by Lind and Bardwell (1923 b) that, in contrast to almost all other materials, most diamonds could not be coloured within a reasonable period by β- and γ-rays from natural sources, but that α-

particle bombardment coloured them green. Prolonged heating at 500°C removed the colour. Later work (Cork 1952 a, b, Hardy 1949, Ehrmann 1950 a, b, Hamilton, Putnam and Ehrmann 1952) established that the green coloration extended only for the range of the bombarding ions. The brown colour occasionally induced is certainly a matter of experimental conditions, and is probably due to a combination of heating and bombardment. It appears that the green colour becomes perceptible when about 10^{-5} of the atoms have been displaced.

Dugdale (1953) used pile irradiation and fast electrons to produce the green coloration, but he also found that with sufficiently low energy electrons (0·5 MeV) when the majority of the defects must have been isolated Frenkel defects, the colour was always blue. This difference was attributed to the effects of the grouping of the defects produced by the higher energy primary knock-ons. Spectroscopic examination of the absorption and fluorescent spectra of irradiated diamond after heat treatment showed that there are very marked similarities to the spectra of some naturally occurring green diamonds. Prolonged heating of the diamonds even at 1000°C did not bleach the colour completely, but turned the diamonds various shades of brown. Pringsheim and Voreck (1952) and Pringsheim (1953) showed that the major increase in the optical absorption coefficient occurs in the near ultra-violet. Heat treatment produces complex effects which have not so far been interpreted satisfactorily. Further spectrophotometric studies have been reported by Ditchburn *et al.* (1955). The diamonds turned blue by low energy electron bombardment do not have the same spectrum as naturally occurring blue diamonds; moreover, heat treatment of the latter produces no change, while the irradiated diamonds turn yellow-green.

(*b*) *Magnetic studies.* Paramagnetic resonance studies of neutron and electron irradiated diamonds (Griffiths *et al.* 1954, 1955) have identified two types of defect. The main effect is an isotropic absorption line whose intensity decreases with heat treatment very much as the coloration does. In addition, an anisotropic spectrum is observed, which is much weaker, but whose intensity is not altered by heating at 1000°C. O'Brien and Pryce (1955) ascribe the isotropic band to isolated Frenkel defects, and the anisotropic absorption spectra to interstitial atoms or vacancies, linked together in pairs, with the axis of the pair parallel to one or other of the [110] directions. For a dose of 10^{19} neutrons cm^{-2} the proportion of displaced atoms was estimated from the intensity of the absorption to be 10^{-3}. This is in adequate agreement with the proportion to be expected from the results of § 2.

In addition to their measurements on quartz, McClelland and Donoghue (1953) observed decreases in the diamagnetic susceptibility of MgO, and none in spinel and BeO; their results on Al_2O_3 were conflicting, and suggest that ionization effects may have contributed to the changes observed.

(*c*) *Luminescence and photoconductivity.* A considerable body of information is available on the effects of radiation on luminescence in scintillators. Marsden (1910) studied decreases of the luminescence of ZnS, $ZnSiO_4$ and $BaPt(CN)_4$, as a function of α-particle bombardment; the luminescence in ZnS was found to be virtually destroyed by doses of about 2×10^{14} α-particles cm^{-2}. Rutherford (1910) interpreted the results in terms of tubes of damage around the track of the α-particles, and obtained tube diameters of 13 Å, 25 Å and 1600 Å in ZnS, $ZnSiO_4$ and $BaPt(CN)_4$ respectively. He assumed that the active centres were destroyed. Recently, Smith and Turkevitch (1954) have reported on the effects

of neutron bombardment on the luminescent efficiency of copper-activated ZnS for doses of up to 10^{17} neutrons cm^{-2}. They conclude that the neutron bombardment produces defects which have a large cross section for trapping excited electrons which then return to their normal levels in non-radiative transitions. A somewhat similar effect, a decrease in the efficiency of α-counting in diamond with neutron bombardment, has been reported recently by Benny and Champion (1954).

Day (1953) studied the photoconductivity induced in magnesium oxide by neutron bombardment, and separated the ionization induced effects from the effects of displaced atoms. The latter showed a saturation, interpreted as 2×10^{-9} defects per atom, setting in at about 10^{15} neutrons cm^{-2}. Such a concentration is too low to account for the 0·1% lattice expansion observed by Primak *et al.* (1953) (see § 5.3.3); moreover, the diamagnetic susceptibility changes of McClelland and Donoghue were still increasing at very much higher doses. It thus seems likely that the saturation arises from the mechanism of photoconductivity rather than from a saturation of the number of interstitial atoms.

5.3.3. *X-ray studies.*

There are available x-ray studies of the effects of room temperature irradiation as a function of dose in zircon ($ZrSiO_4$) (Hurley and Fairbairn 1952, 1953), boron carbide (Tucker and Senio 1954), various boron compounds (Pease 1954), and reports of pile irradiation effects in black phosphorus (Chipman, Warren and Dienes 1953), and various insulators (Primak, Fuchs and Day 1953, Primak 1954). Most of these are concerned with lattice parameter changes and line broadening. Tucker and Senio have examined diffuse scattering, which they find is of the type predicted for solid solutions by Huang (1947) and others; Binnie and Liebschutz (1954) have also examined diffuse scattering, and small angle scattering has been examined in irradiated diamond (Hayes and Smoluchowski 1953, Robinson, Li and Smoluchowski 1954). The accounts so far published of these studies on diffuse and small angle scattering are too sketchy to warrant detailed discussion herein.

(*a*) *Lattice expansions.*

Primak *et al.* report that for a fast neutron dose of about 5×10^{20} neutrons per cm^2, there are no detectable changes in the x-ray photographs of beryl, GeO_2, Ge, Si, corundum (Al_2O_3) and rutile (TiO_2); there are lattice expansions of about 0·1% in MgO, spinel (MgO, Al_2O_3) and CaF_2, and of 1% in diamond and silicon carbide. They also report lattice expansion in Be_2SiO_4 and BeO,Al_2O_3 at a lower dose. Chipman, Warren and Dienes (1953) report a lattice expansion of 0·03% in black phosphorus. They also report the appearance of additional x-ray reflections which seem to be associated with a phase change to the more stable polymorph, red phosphorus. A similar radiation-induced phase change, from white to red phosphorus, is reported by Becquerel (1901). Results obtained by Pease (1954) with powder samples of the compounds BN, B_4C, SrB_6, ZrB_2 and TiB_2 irradiated by means of the ^{10}B (n,α) reaction show lattice parameter changes of the order of 1% at doses of about 10^{18} thermal neutrons cm^{-2}.

It is thus clear that a characteristic effect of the displacement of atoms is an expansion of the lattice, although this may be too small to detect in some cases. For most of the cases reported by Primak *et al.*, one would expect from the results

of § 2 at least 10% of the atoms to have been displaced, and since the lattice expansions could hardly accommodate more than 1% of interstitials, the proportion of interstitials must be at least a factor of ten down on the proportion of atoms displaced. In the case of the compounds showing no detectable effects, recovery must limit the concentration of interstitials even more drastically.

In non-cubic crystals, the lattice expansion is often significantly anisotropic, which can usually be explained, as in quartz, by the nature of the interstitial site available for the displaced atoms. Thus the layer compounds BN and black phosphorus show big lattice expansions perpendicular to the layers. The small contraction in the plane of the layers of BN can be explained as a consequence of puckering of the layers and of the presence of vacancies. The large contraction along the c-axis of B_4C observed by several workers is likewise probably due to a peculiarity of the interstitial sites; it is accompanied by an expansion along the a-axis and a net decrease of density.

Rates of variation of the lattice parameters with dose at zero dose, are given in table 5.3. The doses are expressed as the number of bombarding particles per atom (BPPA). In the case of quartz, the bombarding particles are recoil atoms from fast neutron collisions; in the other cases they are α-particles and recoil nuclei produced by the α-disintegrations.

We also show in table 5.3 whether or not line broadening occurs. Line broadening was detected by Primak *et al.* only in quartz and its polymorphs. In this case, we have seen that the line broadening can be associated with transformation to the metamict state. It seems likely that in the case of BN, and also perhaps of black phosphorus, at low doses, the line broadening can be attributed to the direct effects of interstitial atoms on the lattice periodicity.

Table 5.3. Summary of some X-Ray Studies

Substance	Structure	Initial Expansions $\frac{1}{a}\frac{da}{dx}$[d]	Initial Expansions $\frac{1}{c}\frac{dc}{dx}$[d]	Dose to half Saturation (BPPA)	Line Broadening
BN	Hex.[a]	$-1{\cdot}1$	$+8$	6×10^{-4}	Assoc. with c axis
B_4C	Hex.[b]	$+3{\cdot}5$	-5	13×10^{-4}	Some
SrB_6	Cubic	$+5{\cdot}3$	—	3×10^{-4}	None
TiB_2	Hex.[bc]	$+13$	$+6$	4×10^{-4}	Slight a axis
ZrB_2	Hex.[bc]	$+20$	$+9$	4×10^{-4}	Heavy
$ZrSiO_4$	Tetr.[b]	$+30$[e]		3×10^{-4}	Heavy
α-quartz	Hex.[b]	$+25$	$+8$	Metamict $>3\times10^{-4}$	Heavy

Notes: [a] Layer structure; [b] three dimensional network; [c] isomorphous; [d] the unit of dose, x, is the number of bombarding particles per atom; [e] measured from the (112) reflection only.

(*b*) *Saturation.* In the boron compounds studied by Pease, the lattice expansions were observed to show a marked saturation phenomenon (an example of this has been shown at figure 3.2), which was attributed to a saturation of the number of interstitial atoms. The case of BN was studied in some detail with the aid of saturation theory. The number Q of lattice sites directly adjacent to an interstitial site is twelve, and with this value, saturation is expected to set in at a dose of about 10^{-3} BPPA, with a saturation interstitial concentration of 5%. While saturation was observed to occur at about the right dose, x-ray intensity measurements precluded the expected 5% concentration of interstitials; moreover, irradiation at low temperatures produced much larger effects than at room

temperature. It was therefore concluded that the value of Q was between 100 and 300 at room temperature, and the average number of atoms effectively displaced by a single bombarding particle was only about five. The difference between this last figure and a value of about 50 obtained from Seitz's (1949) expressions was ascribed to virtually instantaneous recovery due to the grouping of the interstitials and vacancies at the end of the ion tracks.

Hurley and Fairbairn (1953) measured the Bragg angle of the 112 reflection from powdered zircons as a function of dose from uranium and thorium impurities, and also obtained a saturation. They fitted an exponential curve to their results and obtained a dose to half-saturation of 3×10^{15} α-particles per milligramme. Primak *et al.* report results for only one dose, so that their results provide no evidence of the presence of saturation.

The doses to half-saturation are given in table 5.3 and, as will be seen, these are all of the same order of magnitude on a scale of bombarding particles per atom. In terms of equation (3.7), this may arise as follows. When Q is large, the number p of displaced atoms per bombarding particle which survive instantaneous recovery is small, because of the grouping of interstitials and vacancies. Likewise when Q is small, p is large. Consequently the saturation dose, $\sim(pQ)^{-1}$, is relatively insensitive to variations of Q from one substance to another. In the case of the zircon results, many more atoms are displaced per bombarding particle, owing to the large contribution from the recoiling nucleus in the α-disintegrations. Because of the great concentration of displaced atoms to be expected in this case, it is possible, however, that very few survive instantaneous recovery.

(*c*) *Non-saturating effects.* In the case of the boron compounds additional effects of some sort or other were observed for doses exceeding the saturation dose; as an example we show in figure 5.12 line breadths measured on x-ray

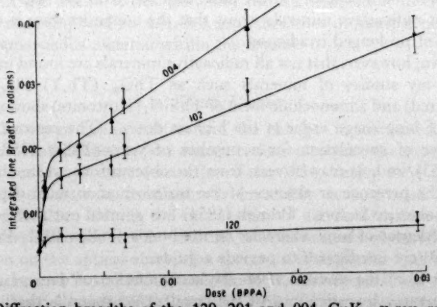


Figure 5.12. Diffraction breadths of the 120, 201 and 004 CuKα x-ray reflections from boron nitride. $\beta(2\theta)$ is defined as $2\int I(\theta)d\theta/I_{\max}$, where θ is the Bragg angle and $I(\theta)$ is the intensity distribution.

photographs of irradiated BN. These non-saturating effects were attributed to two possible causes: the steady destruction of ^{10}B atoms, and the creation of relatively large volumes so heavily disordered that the saturating mechanism can no longer operate. Tucker and Senio (1954) irradiated single crystals of B_4C

and report striking changes in the x-ray patterns which are qualitatively the same at all doses up to 3×10^{20} slow neutrons cm^{-2} (0·2 BPPA), for which dose they report results. This dose is over a hundred times the dose to half-saturation and it therefore appears that Tucker and Senio were measuring non-saturating effects. At this dose, approximately 13% of the boron atoms were destroyed and replaced by 20% of lithium and helium impurities. This would probably be sufficient to account for the effects reported. Further, although each atom of the structure would have been displaced about 10 times on the average, the single crystal x-ray pattern was still observable.

5.3.4. *The metamict state.*

The often observed optical isotropy of metamict minerals was long interpreted as indicating an amorphous state, and this view was confirmed by Vegard's (1916) important observation, subsequently repeated on many other metamict minerals, that metamict $ThSiO_4$ gave no detectable x-ray reflections in spite of a relatively perfect outward morphology. Berzelius (1815) and Wollaston discovered the striking nature of the release of stored energy in gadolinite, in which, by uniform heating of a sample, a sufficient amount of heat can be released suddenly to raise the sample to a high temperature. Faessler's (1942) measurement of the quantity of heat released (89 cal g^{-1}) shows that it was of the same order as the heat of crystallization of such substances as fused quartz. Kulp, Volchok and Holland (1952) have shown by differential thermal analysis that at low doses the stored energy in zircon and samarskite is proportional to dose. While the process of recovery is certainly a recrystallization, this is often extremely complicated, and Pabst (1952) concludes that the original single crystal can only be reconstituted when at least traces of the original lattice are still present. All metamict minerals contain uranium or thorium. These facts, together with the heavy doses we have calculated for radioactive minerals, show that the metamict state is an amorphous end-product of prolonged irradiation.

It is known, however, that not all radioactive minerals are found in the metamict state, and x-ray studies of minerals such as ThO_2, (Th,Y) PO_4,UO_2 (except when weathered) and a monoclinic form of $ThSiO_4$ (Huttonite) show no appreciable destruction of long range order at the highest doses. The reasons for this have been a matter of speculation for a number of years (Pabst 1952, Hurley and Fairbairn 1953), and, as may be seen from the observations on the different forms of $ThSiO_4$, the presence or absence of the transformation must depend critically on relatively obscure factors. Primak (1954) has pointed out that generalizations based on the degree of ionic character in the bonds (Pellas 1951, Primak, Fuchs and Day 1953) are insufficient to provide a guide.

It is clear from the discussion on the accumulation of damage and from the results of the x-ray investigations of the covalent compounds that, provided the crystal structure is maintained, the concentration of single interstitial atoms must saturate, even when there is no thermally activated recovery. For room temperature irradiations, the saturation concentration of interstitial atoms in most cases can hardly be greater than 1% or so, since this is the order of the lattice expansions observed. Thus the displacement of atoms to interstitial sites will not by itself produce the metamict state. Several possible mechanisms may be mentioned by which further disorder can be introduced: (*a*) the crystal structure

becoming unstable at interstitial concentrations of less than the saturation value, (*b*) atoms going to the wrong type of lattice site, (*c*) the disarrangement of relatively large volumes of material in a single event, (*d*) accumulation of impurities. As far as quartz and its polymorphs are concerned, suggestion (*a*) seems the most promising; Wittels and Sherrill (1954) have pointed out that there are grounds for supposing that α-quartz contains ' permanent damage ' at doses exceeding 2×10^{19} neutrons cm^{-2}. Moreover, neither (*b*) nor (*d*) are likely to operate in this case. That naturally irradiated quartz is not metamict may be due to the relatively very slow rate of irradiation, or to the different type of radiation involved. In the case of the boron compounds, it seems likely that (*d*), due to the burn-up of ^{10}B, is mainly responsible for the non-saturating effects; but at least in the case of B_4C, it seems unlikely that this would produce a fully metamict state, even if all the ^{10}B were destroyed.

Hurley and Fairbairn (1953) interpreted the already-mentioned saturation of the changes in the Bragg angle of the 112 reflection from zircon, as the transformation to the metamict state. On the assumption that saturation sets in when all the atoms have been displaced once, they obtained 4500 displaced atoms for each α-disintegration. This number is rather larger than that estimated in § 4.5, which is unusual (see table 2.2), and leaves no margin for recovery of any sort. The density changes represented by the maximum change in Bragg angle is 4%. The density of fully metamict zircon measured by macroscopic methods is about 17% below that of the undamaged (and of fully annealed) material (Anderson and Payne 1937). This discrepancy can hardly be due to experimental errors, and its existence is fully confirmed by the results of earlier workers (Chudoba and von Stackelberg 1936, Lietz 1937). Further, it is difficult to see how an x-ray reflection can be obtained from a *fully* metamict sample. It therefore seems that the saturation of the Bragg angle changes is due to a saturation of the interstitial concentration in the zircon lattice, and that the transformation to the metamict state, which is superimposed on this, is a growth, either in number or in volume, of domains of amorphous material within the crystal.

5.4. *Alkali Halides*

The high density of colour centres and the changes in mechanical properties produced in the alkali halides by low energy radiation require the supposition that ionization introduces crystalline defects (see Seitz 1954, Varley 1954 a, b). Since charged particle or neutron irradiation is invariably accompanied by ionizing radiation, it is difficult to identify the contribution of defects formed by the atomic collision processes discussed in § 2. Studies have been made of the effects of high energy radiation on the optical absorption spectrum (Pringsheim and Yuster 1950, Casler, Pringsheim and Yuster 1950, Przibram 1951, Westervelt 1952), conductivity (Nelson, Sproull and Caswell 1953, Pearlstein 1953, 1954), paramagnetic resonance (Hutchison 1949, Tinkham and Kip 1951), density and mechanical properties (Leivo 1953, Vaughan, Leivo and Smoluchowski 1953, Leivo and Smoluchowski 1954, Westervelt 1954, Binder and Sturm 1954). Satisfactory separation of the effects of ionization from those of collisions has not so far been made.

Acknowledgments

We are grateful to T. M. Fry, R. D. Lowde and J. H. W. Simmons for critical readings of the manuscript, and for helpful suggestions for its correction and improvement. Acknowledgments are also due to those authors and publishers who have given permission for the reproduction of diagrams from their papers, or have allowed us to use unpublished material.

References

Adam, J., and Dugdale, R. A., 1951, *Nature, Lond.*, **168,** 582.
Adam, J., Green, A., and Dugdale, R. A., 1952, *Phil. Mag.*, **43,** 1216.
Allison, S. K., and Warshaw, S. D., 1953, *Rev. Mod. Phys.*, **25,** 779.
Anderson, B. W., and Payne, C. J., 1937, *The Gemmologist*, **7,** 297.
Andrade, E. N. da C., 1945, *Nature, Lond.*, **156,** 113.
Andrew, A., and Davidson, C. R., 1953, *Phys. Rev.*, **89,** 876.
Aronin, L. R., 1954, *J. Appl. Phys.*, **25,** 344.
Barschall, H. H., Rosen, L., Taschek, R. F., and Williams, J. H., 1952, *Rev. Mod. Phys.*, **24,** 1.
Bartlett, J. H., and Dienes, G. J., 1953, *Phys. Rev.*, **89,** 848.
Becquerel, H., 1901, *C.R. Acad. Sci., Paris*, **133,** 709.
Benny, A. H., and Champion, F. C., 1954, *Nature, Lond.*, **173,** 1087.
Berman, R., 1951, *Proc. Roy. Soc.* A, **208,** 90; 1953, *Advances in Phys.* (*Phil. Mag.* Suppl.), **2,** 103.
Berman, R., Foster, E. L., and Rosenberg, H. M., 1955, Report of the Bristol Conference on *Defects in Crystalline Solids* (London: Physical Society).
Berman, R., Klemens, P. G., Simon, F. E., and Fry, T. M., 1950, *Nature, Lond.*, **166,** 864.
Berzelius, J., 1815, *Afhandl. Fys. Kem. Min.*, **4,** 217 (referred to by Pabst 1952).
Bethe, H. A., and Ashkin, J., 1953, *Experimental Nuclear Physics*, **1,** 166, Ed. E. Segrè (New York: Wiley).
Billington, D. S., and Siegel, S., 1950, *Metal Progr.*, **58,** 847.
Binder D., and Sturm, W. J., 1954, *Phys. Rev.*, **96,** 1519.
Binnie, W. P., and Liebschutz, A. M., 1954, *Phys. Rev.*, **94,** 1410.
Blewitt, T. H., and Coltman, R. R., 1951, *Phys. Rev.*, **82,** 769: 1952, *Ibid.*, **85,** 384; 1954, *Acta Metallurgica*, **2,** 549.
Bohr, N., 1948, *K. Danske Vidensk. Selsk. Mat. Fys. Medd.*, **18,** 8.
Bowen, D. B., Eggleston, R. R., and Kropschot, R. H., 1952, *J. Appl. Phys.*, **23,** 630.
Bowen, D., and Rodeback, G. W., 1953, *Acta Metallurgica*, **1,** 649.
Bradley, R. C., 1954, *Phys. Rev.*, **93,** 719.
Brattain, W. H., and Pearson, G. L., 1950, *Phys. Rev.*, **80,** 846.
Briggs, G. H., 1954, *Rev. Mod. Phys.*, **26,** 1.
Brinkman, J. A., 1954, *J. Appl. Phys.*, **25,** 961.
Brinkman, J. A., Dixon, C. E., and Meechan, C. J., 1954, *Acta Metallurgica*, **2,** 38.
Broegger, W. C., 1893, *Salmonsens store illustrededе Konversationslexicon*, **1,** 742. (Abstract in *Z. Kristallogr.*, 1896, **25,** 427)
Broom, T., 1954, *Advances in Physics*, (*Phil. Mag.* Suppl.) **3,** 26.
Brown, C. S., and Thomas, L. A., 1952, *Nature, Lond.*, **169,** 35.
Brown, W. L., Fletcher, R. C., and Wright, K. A., 1953, *Phys. Rev.*, **92,** 591.
Burton, M., 1947, *J. Phys. Chem.*, **51,** 611.
Casler, R., Pringsheim, P. and Yuster, P., 1950, *J. Chem. Phys.*, **18,** 887.
Charlesby, A., 1954, *Nucleonics*, **12,** No. 6, 18.
Chipman, D. L., Warren, B. E., and Dienes, G. J., 1953, *J. Appl. Phys.*, **24,** 1251.
Choong, S. P., 1945, *Proc. Phys. Soc.*, **57,** 49.
Chudoba, K., and von Stackelberg, M., 1936, *Z. Kristallogr.*, **95,** 230.
Cleland, J. W., Billington, D. S., and Crawford, J. H., 1953, *Phys. Rev.*, **91,** 238.
Cleland, J. W., and Crawford, J. H., 1954a, *Phys. Rev.*, **93,** 894; 1954b, *Ibid.*, **94,** 1410; 1954c, *Ibid.*, **95,** 1177.

CLELAND, J. W., CRAWFORD, J. H., LARK-HOROVITZ, K., PIGG, J. C., and YOUNG, F. W., 1951a, *Phys. Rev.*, **83,** 312; 1951b, *Ibid.*, **84,** 861.
CLELAND, J. W., LARK-HOROVITZ, K., and PIGG, J. C., 1950, *Phys. Rev.*, **78,** 814.
COLTMAN, R. R., and BLEWITT, T. H., 1953, *Phys. Rev.*, **91,** 236.
COOK, L. G., and CUSHING, R. L., 1953, *Acta Metallurgica*, **1,** 539.
COOPER, H. G., KOEHLER, J. S., and MARX, J. W., 1954, *Phys. Rev.*, **94,** 496.
CORK, J. M., 1942a, *Phys. Rev.*, **62,** 80; 1942b, *Ibid.*, **6,** 318.
CRAWFORD, J. H., CLELAND, J. W., HOLMES, D. K., and PIGG, J. C., 1953, *Phys. Rev.*, **91,** 243.
CRAWFORD, J. H., and LARK-HOROVITZ, K., 1950, *Phys. Rev.*, **78,** 815.
DAINTON, F. S., and ROWBOTTOM, J., 1954, *Trans. Faraday Soc.*, **50,** 480.
DAVIS, R. E., JOHNSON, W. E., LARK-HOROVITZ, K., and SIEGEL, S., 1948, *Phys. Rev.*, **74,** 1255.
DAY, H. R., 1953, *Phys. Rev.*, **91,** 822.
DENNEY, J., 1953, *Phys. Rev.*, **92,** 531; 1954a, *U.S. AEC Report NAA–SR–*271; 1954b, *Phys. Rev.*, **94,** 1417.
DEXTER, D. L., 1952, *Phys. Rev.*, **87,** 768.
DIECKAMP, H., and CRITTENDEN, E. C., 1954, *Phys. Rev.*, **94,** 1418.
DIENES, G. J., 1952a, *Phys. Rev.*, **86,** 228; 1952b, *Ibid.*, **87,** 666; 1953a, *Ann. Rev. Nuclear Sci.*, **2,** 187; 1953b, *J. Appl. Phys.*, **29,** 666; 1953c, *Phys. Rev.*, **91,** 1283.
DITCHBURN, R. W., MITCHELL, E. W. J., PAIGE, E. G. S., CUSTERS, J. F., DYER, H. B., and CLARK, C. D., 1955, Report of Bristol Conference on *Defects in Crystalline Solids* (London: Physical Society).
DIXON, C. E., and BOWEN, D. B., 1954, *Phys. Rev.*, **94,** 1418.
DIXON, C. E., and MEECHAN, C. J., 1953, *Phys. Rev.*, **91,** 237.
DIXON, C. E., MEECHAN, C. J., and BRINKMAN, J. A., 1953, *Phil. Mag.*, **44,** 449.
DUGDALE, R. A., 1952, *Phil. Mag.*, **43,** 912; 1953, *Brit. J. Appl. Phys.*, **4,** 337; 1955, Report of Bristol Conference on *Defects in Crystalline Solids* (London: Physical Society).
DUGDALE, R. A., and GREEN, A., 1954, *Phil. Mag.*, **45,** 163.
EGGEN, D. T., and LAUBENSTEIN, M. J., 1953, *Phys. Rev.*, **91,** 238.
EGGLESTON, R. R., 1953, *Acta Metallurgica*, **1,** 679.
EGGLESTON, R. R., and BOWMAN, F. E., 1953, *J. Appl. Phys.*, **24,** 229.
EHRMANN, M. L., 1950a, *Gems*, **6,** 295; 1950b, *Ibid.*, **6,** 318.
FAESSLER, A., 1942, *Z. Kristallogr.*, **104,** 81.
FAN, H. Y., and LARK-HOROVITZ, K., 1955, Report of the Bristol Conference on *Defects in Crystalline Solids* (London: Physical Society).
FILLNOW, R. H., HALTEMAN, E. K., and MECHLIN, G. F., 1953, *Phys. Rev.*, **91,** 236.
FLEEMAN, J., 1953, *Phys. Rev.*, **91,** 237; 1954, *Ibid.*, **94,** 1422.
FLETCHER, R. C., and BROWN, W. L., 1953, *Phys. Rev.*, **92,** 585.
FLORIDA, C. D., HOLT, F. R., and STEPHEN, J. H., 1954, *Nature, Lond.*, **173,** 397.
FORMAN, G., 1949, *Phys. Rev.*, **76,** 196.
FORSTER, J. H., FAN, H. Y., and LARK-HOROVITZ, K., 1952, *Phys. Rev.*, **86,** 643; 1953, *Ibid.*, **91,** 229.
GEIB, I. G., and GRACE, R. E., 1952, *Phys. Rev.*, **86,** 643.
GLICK, H. L., BROOKS, F. C., WITZIG, W. F., and JOHNSON, W. E., 1952, *Phys. Rev.*, **87,** 1074.
GLICK, H. L., and WITZIG, W. F., 1953, *Phys. Rev.*, **91,** 236.
GRIFFITHS, J. H. E., OWEN, J., and WARD, J. M., 1954, *Nature, Lond.*, **173,** 439; 1955, Report of Bristol Conference on *Defects in Crystalline Solids* (London: Physical Society).
GROVE, W. R., 1852, *Phil. Trans. Roy. Soc.*, **142,** 87.
HAHN, O., 1936, *Applied Radiochemistry* (Ithaca: Cornell U.P.).
HAMBERG, A., 1914, *Geol. Fören. Stockh. Förh.*, **36,** 31 (referred to by Pabst, 1952).
HAMILTON, J. G., PUTNAM, T. M., and EHRMANN, M. L., 1952, *Amer. Min.*, **37,** 941.
HARDY, J. A., 1949, *Gems*, **6,** 167.
HAYES, S., and SMOLUCHOWSKI, R., 1953, *Phys. Rev.*, **91,** 244.
HUANG, K., 1947, *Proc. Roy. Soc.* A, **190,** 102.
HUFFMAN, J. R., 1954, *Nucleonics*, **12,** No. 4, 20.
HUNTINGTON, H. B., 1942, *Phys. Rev.*, **61,** 325; 1953, *Ibid.*, **91,** 1092; 1954, *Ibid.*, **93,** 1414.
HUNTINGTON, H. B., and SEITZ, F., 1942, *Phys. Rev.*, **61,** 315.

HURLEY, P. M., and FAIRBAIRN, H. W., 1952, *J. Appl. Phys.*, **23,** 1408; 1953, *Bull. Geol. Soc. Amer.*, **64,** 659.
HUTCHISON, C. A., 1949, *Phys. Rev.*, **75,** 1769.
JAMES, H. M., and LARK-HOROVITZ, K., 1951, *Z. phys. Chem.*, **198,** 107.
JOHNSON, F. B., and PEASE, R. S., 1954, *Phil. Mag.*, **45,** 651.
JOHNSON, R. D., and MARTIN, A. B., 1952, *J. Appl. Phys.*, **23,** 1245.
JOHNSON, W. E., and LARK-HOROVITZ, K., 1949, *Phys. Rev.*, **76,** 442.
JOLY, J., 1907, *Phil. Mag.*, **13,** 309.
JONES, E. R. W., MUNRO, W., and HANCOCK, N. H., 1954, *J. Nucl. Energy*, **1,** 76.
JONGENBURGER, P., 1953a, *Phys. Rev.*, **90,** 710; 1953b, *Appl. Sci. Res., Hague* B, **3,** 237.
KEESOM, P. H., LARK-HOROVITZ, K., and PEARLMAN, N., 1952, *Science*, **116,** 630.
KEYWELL, F., 1952, *Phys. Rev.*, **87,** 161.
KIERSTEAD, H. A., 1954, *Bull. Amer. Phys. Soc.*, **29,** 30.
KINCHIN, G. H., 1953, *Proc. Roy. Soc.* A, **217,** 9; 1954, *J. Nucl. Energy*, **1,** 124.
KINCHIN, G. H., and PEASE, R. S., 1955, *J. Nucl. Energy*, **1,**
KITTEL, C., 1953, *Introduction to Solid State Physics* (New York: Wiley).
KLEMENS, P. G., 1951, *Proc. Roy. Soc.* A, **208,** 108.
KLONTZ, E. E., 1952, *U.S. AEC Report*, AECU 2267.
KLONTZ, E. E., and LARK-HOROVITZ, K., 1951, *Phys. Rev.*, **82,** 763; 1952, *Ibid.*, **86,** 643.
KNIPP, J., and TELLER, E., 1941, *Phys. Rev.*, **59,** 659.
KOEHLER, J. S., and SEITZ, F., 1954, *Z. Phys.*, **138,** 238.
KUBASCHEWSKI, O., 1950, *Trans. Faraday Soc.*, **46,** 713.
KULP, J. L., VOLCHOK, H. L., and HOLLAND, H. D., 1952, *Amer. Min.*, **37,** 709.
KUNZ, F. W., and HOLDEN, A. N., 1954, *Acta Metallurgica*, **2,** 816.
LARK-HOROVITZ, K., 1951, *Semi-Conducting Materials*, Ed. H. K. HENISCH (London: Butterworths Scientific Publications), p. **47.**
LARK-HOROVITZ, K., BECKER, M., DAVIS, R. E., and FAN, H. Y., 1950, *Phys. Rev.*, **78,** 334.
LARK-HOROVITZ, K., BLEULER, E., DAVIS, R. E., and TENDAM, D., 1948, *Phys. Rev.*, **73,** 1256.
LEIVO, W. J., 1953, *Phys. Rev.*, **91,** 245.
LEIVO, W. J., and SMOLUCHOWSKI, R., 1954, *Phys. Rev.*, **93,** 1415.
LIETZ, J., 1937, *Z. Kristallogr.*, **98,** 201.
LIND, S. C., 1928, *Chemical Effects of Alpha Particles and Electrons* (New York: Chemical Catalog Co.)
LIND, S. C., and BARDWELL, D. C., 1923a, *J. Franklin Inst.*, **196,** 375; 1923b, *Ibid.*, **196,** 521.
MAKIN, M. J., 1955, *J. Nucl. Energy*, **1.**
MARTIN, A. B., AUSTERMAN, S. B., EGGLESTON, R. R., MCGEE, J. F., and TARPINIAN, M., 1951, *Phys. Rev.*, **81,** 664.
MARSDEN, E., 1910, *Proc. Roy. Soc.* A, **83,** 548.
MARX, J. W., COOPER, H. G., and HENDERSON, J. W., 1952, *Phys. Rev.*, **88,** 106.
MASSEY, H. S. W., and BURHOP, E. H. S., 1952, *Electronic and Ionic Impact Phenomena* (Oxford: University Press).
MAYBURG, S., 1954, *Phys. Rev.*, **95,** 38.
MCCLELLAND, J. D., and DONOGHUE, J. J., 1953, *J. Appl. Phys.*, **24,** 963.
MCDONELL, W. R., and KIERSTEAD, H. A., 1954, *Phys. Rev.*, **93,** 247.
MCKAY, H. A. C., 1950, *Prog. Nuclear Phys.*, **1,** 168.
MCKINLEY, W. A., and FESHBACH, H., 1948, *Phys. Rev.*, **74,** 1759.
MCREYNOLDS, A. W., AUGUSTYNIAK, W., MCKEOWN, W., and ROSENBLATT, D. B., 1954, *Phys. Rev.*, **94,** 1417.
MEECHAN, C. J., and EGGLESTON, R. R., 1954, *Phys. Rev.*, **93,** 953.
MEYER, R. A., 1954, *J. Appl. Phys.*, **25,** 1369.
MITCHELL, E. W. J., and PAIGE, E. G. S., 1954, *Proc. Phys. Soc.*, B, **67,** 262.
MOSS, T. S., 1952, *Photoconductivity in the Elements* (London: Butterworths Scientific Publications), p. 105.
MOTT, N. F., 1929, *Proc. Roy. Soc.* A, **124,** 425; 1932, *Ibid.*, **135,** 429.
MURRAY, G. T., and TAYLOR, W. E., 1954, *Acta Metallurgica*, **2,** 52.
NABARRO, F. R. N., 1952, *Phys. Rev.*, **87,** 665.
NELSON, C. M., SPROULL, R. L., and CASWELL, R. S., 1953, *Phys. Rev.* **90,** 364.
NIX, F. C., and SHOCKLEY, W., 1938, *Rev. Mod. Phys.*, **10,** 1.

O'Brien, M. C. M., and Pryce, M. H. L., 1955, Report of Bristol Conference on *Defects in Crystalline Solids* (London: Physical Society).
Overhauser, A. W., 1953, *Phys. Rev.*, **90,** 393; 1954, *Ibid.*, **94,** 1551.
Ozeroff, J., 1949, *U.S. AEC Report* AECD 2973.
Pabst, A., 1952, *Amer. Min.*, **37,** 137.
Pearlstein, E., 1953, *Phys. Rev.*, **92,** 881; 1954, *Ibid.*, **94,** 1409.
Pease, R. S., 1954, *Acta Cryst., Camb.*, **7,** 633.
Pellas, P., 1951, *C.R. Acad. Sci., Paris,* **233,** 1369.
Pepper, R., Klontz, E., Lark-Horovitz, K., and MacKay, J., 1954, *Phys. Rev.*, **94,** 1410.
Primak, W., 1954, *Phys. Rev.*, **95,** 837.
Primak, W., Fuchs, L. H., and Day, P., 1953, *Phys. Rev.*, **92,** 1064.
Pringsheim, P., 1953, *Phys. Rev.*, **91,** 551.
Pringsheim, P., and Voreck, R. S., 1952, *Z. Phys.*, **133,** 2.
Pringsheim, P., and Yuster, P., 1950, *Phys. Rev.*, **78,** 293.
Przibram, K., 1951, *Z. Phys.*, **130,** 269.
Redman, J. K., Coltman, R. R., and Blewitt, T. H., 1953, *Phys. Rev.*, **91,** 448.
Randolph, B., 1952, *Phys. Rev.*, **85,** 710.
Robinson, W. H., Li, Y. Y., and Smoluchowski, R., 1954, *Phys. Rev.*, **94,** 1435.
Rosenblatt, D. B., Smoluchowski, R., and Dienes, G. J., 1954, *Phys. Rev.*, **94,** 1417.
Rutherford, E., 1910, *Proc. Roy. Soc.* A, **83,** 561.
Rutherford, E., Chadwick, J., and Ellis, C. D., 1930, *Radiations from Radioactive Substances* (Cambridge: University Press).
Schneider, E. G., and O'Bryan, H. M., 1937, *Phys. Rev.*, **51,** 293.
Seitz, F., 1949, *Disc. Faraday Soc.*, **5,** 271; 1952a, *Phys. To-day*, **5,** 6; 1952b, *Advances in Phys.*, **1,** 43; 1954, *Rev. Mod. Phys.*, **26,** 7.
Sidhu, S. S., and Henry, C. O., 1950, *Phys. Rev.*, **80,** 123.
Siegel, S., 1949, *Phys. Rev.*, **75,** 1823.
Sisman, O., and Bopp, C. D., 1951, *U.S. AEC Report* ORNL 928.
Slater, J. C., 1951, *J. Appl. Phys.*, **22,** 237.
Smith, A. W., and Turkevitch, J., 1954, *Phys. Rev.*, **94,** 859.
Snyder, W. S., and Neufeld, J., 1954, *Phys. Rev.*, **94,** 760.
Stevens, D. K., Cleland, J. W., and Crawford, J. H., 1954, *Phys. Rev.*, **94,** 1409.
Taylor, B. T., 1952, *AERE Report No. N/R.* 1005.
Tinkham, M., and Kip, A. F., 1951, *Phys. Rev.*, **83,** 657.
Townes, C. H., 1944, *Phys. Rev.*, **65,** 319.
Tucker, C. W., and Sampson, J. B., 1954, *Acta Metallurgica*, **2,** 433.
Tucker, C. W., and Senio, P., 1954, *Acta Cryst., Camb.*, **7,** 456.
Varley, J. H. O., 1954a, *Nature, Lond.* **174,** 886; 1954b, *J. of Nucl. Energy*, **1,** 130.
Vaughan, W. H., Leivo, W. J., and Smoluchowski, R., 1953, *Phys. Rev.*, **91,** 245.
Vegard, L., 1916, *Phil. Mag.*, **32,** 65.
Watt, B. E., 1952, *Phys. Rev.*, **87,** 1037.
Wells, H. G., 1909, *Tono-Bungay* (London: Macmillan).
Werner, G. K., 1954, *Phys. Rev.*, **93,** 633.
Westervelt, D., 1952, *Phys. Rev.*, **86,** 643; 1954, *U.S. AEC Report* NAA–SR–888.
Wigner, E. P., 1946, *J. Appl. Phys.*, **17,** 857.
Williams, E. J., 1945, *Rev. Mod. Phys.*, **17,** 217.
Wittels, M., 1953, *Phys. Rev.*, **89,** 657.
Wittels, M., and Sherrill, F. A., 1954, *Phys. Rev.*, **93,** 1117.
Witzig, W. F., 1952, *J. Appl. Phys.*, **23,** 1263.
Wruck, D., and Wert, C., 1954, *Phys. Rev.*, **94,** 1417.
Yockey, H. P., Andrew, A., Fillmore, F. L., Glasgow, L. E., Hunt, C. d'A., and Pepper, J. H., 1954, *Rev. Sci. Instrum.*, **25,** 1011.

IONIZATION BY RELATIVISTIC PARTICLES

By B. T. PRICE

Atomic Energy Research Establishment, Harwell, Didcot, Berks.

CONTENTS

	PAGE
§ 1. Introduction	52
§ 2. The ionization process	54
2.1. The average energy-loss by collision	54
2.2. The primary ionization and the energy-loss per ion pair	57
2.3. Fluctuations of energy-loss; the most probable energy-loss	58
§ 3. The density effect	60
§ 4. Results obtained with gaseous detectors	64
4.1. Proportional counters and pressure ionization chambers	64
4.2. Cloud chambers	69
4.3. Low pressure Geiger counters	72
§ 5. Results obtained with condensed detectors	74
5.1. Photographic plates	74
5.2. Crystal counters and scintillators	78
5.3. Energy-loss of electrons in thin foils	78
Acknowledgments	79
References	79

Abstract. A summary is given of the theories of the relativistic increase of energy-loss by ionization and of the density effect. The experimental evidence is then discussed and compared with theoretical predictions in a number of sections, each dealing with a particular experimental technique: proportional counters, cloud chambers, low pressure Geiger counters, photographic plates, crystal counters and scintillators. A brief description is also given of experiments in which the energy-loss of fast electrons in thin foils is measured.

§ 1. Introduction

During recent years a considerable effort has been devoted to the study of the energy-loss by ionization of fast charged particles. This is a field of work in which the theory was largely established well before any experimental verification was possible, and in which spectacular discoveries were neither expected nor obtained. The particular interest of this work lies rather in the large measure of detailed agreement that now exists between theory and experiment, in a field where precise experimentation is often extremely difficult.

According to the theories developed by Bohr (1915, 1948), Möller (1932), Bethe (1932, 1933), and Bloch (1933) the most probable energy-loss per unit path length (Δ_0) of a fast charged particle passing through matter should have a minimum value when the velocity of the particle $v \simeq 0{\cdot}96\,c$. At lower velocities Δ_0 should vary as $1/v^2$, while for a relativistic particle the energy-loss should increase logarithmically with the energy of the particle (this behaviour is illustrated in figure 1). Further, Δ_0 should depend only on the velocity and charge, and not on the mass μ of the particle. The 'logarithmic increase' (i.e. with energy) arises essentially from a relativistic increase in the radius of action of the electric field of the particle as its velocity approaches the velocity of light. It was pointed

out by Swann (1938) that, in the case of condensed media, polarization effects should screen the more distant atoms from the field of the particle. The most probable energy-loss should therefore not increase indefinitely but should saturate at some particle velocity determined by the polarization properties of the medium. A simplified theoretical treatment given by Fermi (1939, 1940) showed that the polarization effects should depend principally on the electron density in the

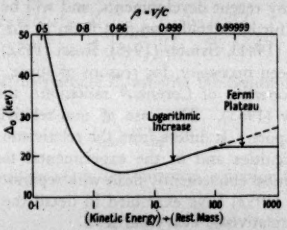


Figure 1. Dependence of the most probable energy-loss Δ_0 of a singly charged particle on its velocity v and energy. The curve is drawn for a 10 cm thickness of oxygen at S.T.P.

medium. They should be small for gases, except at extremely high energies ($E \sim 100\ \mu c^2$), but important and easy to verify experimentally for condensed media (figures 8, 9, 10, 13). For this reason the saturation of the rate of energy-loss is usually referred to as the 'density effect'. Although very considerable elaboration of Fermi's arguments has since been undertaken by many authors, these general conclusions have not been challenged. There has, however, been much discussion concerning the detailed theory of the density effect, and the exact amount of energy which is dissipated in the form of Cerenkov radiation.

It has only recently become possible to place the experimental verification of the theory on anything like a firm footing. Before 1939 almost the only ways of measuring the ionization of fast particles (usually cosmic rays) were by drop counting in a cloud chamber and by knowing the efficiency of a low-pressure Geiger counter. Today we have in addition not only gas-filled proportional counters and pulse ionization chambers of high accuracy, but also scintillation and crystal counters and electron-sensitive photographic plates, which make possible the study of ionization in condensed media. A further difficulty in the early attempts at verification of the 'logarithmic increase' for cosmic-ray particles was the very slow variation of ionization with energy; this meant that the measurements had to be extended to extremely high energies in order to obtain an appreciable increase above the minimum value of the ionization. The difficulty of measuring such energies by magnetic deflection techniques for a time limited the investigations to the case of electrons, which for a given value of $\beta\,(=v/c)$ (and therefore of ionization) are much more easily deflected than mesons. However, at least one instrument—the Manchester spectrometer (Hyams *et al.* 1950)—is now capable of measuring meson energies of 10^{12} ev (or 10^4 rest-masses), so that for mesons at least the whole range of relativistic ionization phenomena, including the onset of the density effect in gases, can now be investigated. The case of relativistic protons and heavier particles is one stage more difficult, and has not yet been given much attention. Finally, the third of the

pre-war difficulties, that of intensity, is being rapidly overcome as accelerating machines of higher and higher energy become available.

The purpose of this Progress Report is to summarize the best of the experimental evidence (most of which has been obtained during the last four years) and to indicate the lines on which current theories of the density effect have been developed. The general theory of ionization and of the logarithmic rise itself have not undergone any recent developments, and will be discussed very briefly; readers who wish for further details can find them in the comprehensive surveys by Rossi and Greisen (1941), Symon (1948), Rossi (1952) and Bethe and Ashkin (1952). It has also been necessary, for reasons of space, to assume a knowledge of the elementary properties of Cerenkov radiation; these have recently been summarized by Jelley (1953). The case of non-relativistic particles has been excluded from this Report. It differs from the relativistic case both in the nature of its theoretical difficulties and in the experimental techniques employed, so that the two cases are most conveniently dealt with separately. In a recent review, Allison and Warshaw (1953) have examined in detail the published experiments on ionization by non-relativistic heavy particles.

§ 2. The Ionization Process

2.1. *The Average Energy-Loss by Collision*

(*a*) *Particles heavier than electrons.*

When a heavy, fast, charged particle passes through matter, it loses energy principally by inelastic collisions with atomic electrons, so causing excitation and ionization. An individual atomic electron may receive in a single collision any amount of energy from $\epsilon_{\min} = I$ to $\epsilon_{\max} = T$, depending on the impact parameter for the particular collision. The lower limit I (which is usually referred to as the mean excitation or ionization potential of the atom) is a measure of the smallest amount of energy which can be transferred, on the average, to a bound electron. It is difficult to estimate theoretically, and is best determined by experiment. For instance, the measurements of Bakker and Segrè (1951), and of Mather and Segrè (1951) show that I/Z is a function of Z varying smoothly from about 15 ev for the lightest elements to 9·6 ev for the heaviest; the approximation $I = I_0 Z \simeq 13{\cdot}5\ Z$ ev (which has been frequently employed in the literature since its use was suggested by Bloch (1933)) is usually sufficiently accurate, since it is $\log I$, and not I itself, which is important in the theory.

The upper limit T is set by the maximum energy that can be transferred to an electron in a single collision. For a particle of mass μ electron masses ($\mu > 1$) and total energy E_0 Bhabha (1937) has shown that

$$\epsilon_{\max} = T = \frac{E_0^2 - \mu^2 c^4}{\mu c^2(\mu/2m + m/2\mu + E_0/\mu c^2)} \qquad \ldots\ldots(1)$$

where m is the electronic mass.

The probability $W(\epsilon)d\epsilon$ of an electron (considered as unbound) receiving an amount of energy between ϵ and $\epsilon + d\epsilon$ from an energetic but non-relativistic particle is given by the classical Rutherford formula (Livingstone and Bethe 1937)

$$W(\epsilon)d\epsilon = \frac{2\pi N e^4}{mv^2} \frac{\rho Z z^2}{A} \frac{d\epsilon}{\epsilon^2} \qquad \ldots\ldots(2)$$

where e is the charge on the electron, N is Avogadro's number, ρ the density of

the medium of atomic number Z and atomic weight A, and v the velocity of the particle, which carries a charge z. The quantity $N\rho Z/A$ is simply the number of electrons per unit volume.

The average energy-loss per unit path length dE/dx resulting from the combined effect of all such elementary collisions in the relativistic case can then be shown by a quantum-mechanical treatment (see Rossi 1952) to be

$$\frac{dE}{dx} = \frac{2\pi z^2 N e^4 \rho}{mv^2} \frac{Z}{A} \left[\ln \left\{ \frac{2m c^2 \beta^2 T}{(1-\beta^2) I_0^2 Z^2} \right\} - 2\beta^2 \right] \quad \ldots\ldots(3)$$

provided the energy of the particle E_0 satisfies the condition

$$\frac{E_0}{\mu c^2} \ll \frac{\mu}{m}. \quad \ldots\ldots(4)$$

If $\mu \gg m$, and condition (4) is also satisfied, equation (1) reduces to

$$T \simeq 2mc^2 \frac{\beta^2}{(1-\beta^2)} \quad \ldots\ldots(5)$$

and the expression for the average rate of energy loss becomes

$$\frac{dE}{dx} = \frac{4\pi z^2 N e^4 \rho}{mv^2} \frac{Z}{A} \left[\ln(2mv^2/I_0 Z) - \ln(1-\beta^2) - \beta^2 \right]. \quad \ldots\ldots(6)$$

In some experiments it is necessary to ignore all collisions in which the energy transferred is greater than some value T_{exp}, which may be very much smaller than the value given by (5). Theory and experiment are then compared by putting the appropriate value for T in (3). If $T_{exp} \sim 1$ kev, as is usual in cloud chamber experiments, then dE/dx approximates to the most probable energy-loss given by (11), and is indeed often loosely referred to as the probable energy-loss.

The $\ln(1-\beta^2)$ term leads to the interesting consequence, which particularly concerns us here, that the energy-loss by ionization and excitation of relativistic particles ($\beta > 0{\cdot}97$) increases logarithmically with the momentum $p = \mu c \beta(1-\beta^2)^{-1/2}$. Part of this increase is due to the increase of the maximum transferable energy T with E_0, and part to the effect of the Lorentz contraction in adding to the range of the electric forces of the ionizing particle. At lower energies the average energy-loss per unit length dE/dx increases rapidly as β decreases, as shown in figure 1.

Formula (6) ceases to be strictly valid at very high energies, when condition (4) is no longer fulfilled. It is then necessary to take into account the spin of the particle which modifies the collision probability for very close collisions and thus affects the average energy-loss (Rossi and Greisen 1941, Bethe and Ashkin 1952). However, for reasons given later most experiments measure not the average but the most probable ionization, which, being determined almost entirely by the distant collisions, will not be affected by spin.

(*b*) *Electrons.*

The theory of energy-loss of (negative) electrons differs in two important respects from that just given. First, the incident and struck particles are indistinguishable, so that the expression for T becomes $\frac{1}{2}(E_0 - mc^2)$ in place of that

given in (1). The complete formula for the average energy-loss by collision was given by Møller (1932); it reduces in the extreme relativistic case to

$$\frac{dE}{dx} = \frac{4\pi N e^4 \rho}{mc^2} \frac{Z}{A} \left[\ln (2mv^2/I_0 Z) - \tfrac{3}{4} \ln(1-\beta^2) - \tfrac{1}{2} \ln 8 + \tfrac{1}{16} \right] \quad \ldots\ldots(7)$$

a formula which gives a somewhat different dependence on β from that given by formula (6).

The second difference between electrons and heavy particles is that when electrons are accelerated in the field of a nucleus or of a charged particle, they radiate energy in the form of quanta ('bremsstrahlung'); for heavier particles the effect is negligible. This mode of losing energy becomes increasingly important as the energy of the electron and the atomic number of the absorber are increased. At some critical energy E_c, the ionization and radiation losses will be equal. For high energies electrons have their energy reduced on the average by a factor e after travelling a distance known as one radiation length (X_0). In travelling this distance through matter they radiate on the average several quanta, in one of which a considerable fraction of their energy is likely to be lost. The accompanying table taken from Bethe and Ashkin (1952) gives values of E_c and X_0 for various substances.

Table 1. Radiation Lengths and Critical Energies

Substance	Z	X_0(g cm^{-2})	E_c(Mev)
Hydrogen	1	58	340
Helium	2	85	220
Carbon	6	42·5	103
Nitrogen	7	38	87
Oxygen	8	34·2	77
Aluminium	13	23·9	47
Argon	18	19·4	34·5
Iron	26	13·8	24
Copper	29	12·8	21·5
Lead	82	5·8	6·9
Air	—	36·5	83
Water	—	35·9	93

According to an approximate formula of Bethe and Heitler (1934)

$$\frac{E_c}{mc^2} \simeq \frac{1600}{Z} \quad \ldots\ldots(8)$$

and

$$\frac{\text{Radiation loss}}{\text{Collision loss}} \simeq \frac{E_0}{E_c}. \quad \ldots\ldots(9)$$

The emission of radiation severely limits the range of experiments which attempt to measure the *energy-loss by collision* of electrons in passing through appreciable thicknesses of materials of high atomic number. On the other hand, the problems of measuring the *ionization* of electrons are scarcely different from the case of heavy particles, since the useful depth of a counter or photographic plate is usually small compared with one radiation length. Even if a photon were to be emitted, it would probably escape from the detector. The effect on the experiment would then be a reduction in the energy of the electron; but the

slow variation of ionization with energy tends to minimize the error introduced by such an occurrence.

2.2. *The Primary Ionization and the Energy-Loss per Ion Pair*

Many experiments measure the ionization (but not the excitation) liberated by the passage of a charged particle, rather than the energy-loss discussed above. The primary specific ionization in a gas J_{pr} is defined as the average number of ionizing collisions per centimetre at S.T.P. According to Bethe (1933)

$$J_{pr} = 2\pi r_0^2 mc^2 n_0 z^2 \frac{Z}{I\beta^2}\left[\ln\left(\frac{2mc^2\beta^2}{(1-\beta^2)I_0 Z}\right) + b - \beta^2\right] \qquad \ldots\ldots(10)$$

where $r_0 = e^2/mc^2$ is the classical electron radius, n_0 the number of atoms per cm^3 of gas and a and b are two dimensionless constants depending on the electronic structure of the gas. Bethe was able to calculate a and b for *atomic* hydrogen, but in other cases the calculation becomes difficult. An experimental determination of these constants can be made by fitting formula (10) to a set of experimental points, such as those shown in figure 12 (McClure 1953). An explicit expression for J_{pr} in hydrogen has recently been given by Budini and Taffara (1953).

A fraction of the primary electrons ejected by direct interaction with the ionizing particle will be themselves sufficiently energetic to cause further (secondary) ionization; the total ionization (primary plus secondary) is rather more than double the primary ionization. The ratio of the mean energy loss to the total number of ion pairs produced—i.e. the mean energy taken to produce an ion pair—is found for most substances, and especially the rare gases, to be practically independent of the nature or energy of the ionizing particle. This fortunate accident of nature places a powerful tool in the hands of the experimentalist: it enables him to replace the difficult measurement of energy-loss by the comparatively straightforward measurement of ionization.

It is usual to assume that the energy required to form an ion pair is exactly constant for a given material; for instance, this assumption underlies much of the work described in this Report. The extent to which one is justified in making such an assumption is discussed in detail by Bethe and Ashkin (1952), Cranshaw (1952) and West (1953 b). To illustrate the constancy of the energy Q required to form an ion pair, some experimental results of Valentine and Curran (1952) and Bakker and Segrè (1951) are reproduced in table 2; these values are thought to be correct to 2%. As far as is known, for all fast particles Q is constant to within a few per cent, but the evidence is still incomplete and further measurements of high accuracy are certainly needed.

Table 2. Energy Q (ev) required to form one Ion Pair

Gas	Slow electrons (5 kev)		Polonium alphas		340 Mev protons
	Q	R	Q	R	R
Argon	27·0	1·00	25·9	1·00	1·00
Helium	32·5	1·20	31·7	1·22	1·17
Hydrogen	38·0	1·41	37·0	1·43	1·38
Nitrogen	35·8	1·32	36·0	1·39	1·31
Air	35·0	1·29	35·2	1·36	1·30
Oxygen	32·2	1·19	32·2	1·24	1·23
Methane	30·2	1·11	29·0	1·12	—

$R = Q_{gas}/Q_{argon}$

2.3. *Fluctuations of Energy-Loss; the most probable Energy-Loss*

Experimental determinations of the energy-loss of a fast particle frequently involve the measurement of the ionization liberated in 'thin' detectors, i.e. detectors in which the particle suffers a considerable number of ionizing collisions, but does not lose more than a small fraction of its original energy. Under these conditions there are very marked fluctuations in the observed rate of energy-loss (Bohr 1915, Williams 1929, Landau 1944, Symon 1948, Blunck *et al.* 1950, 1951). These fluctuations arise partly from the Poisson distribution of ionizing events, and partly from the wide range of energies which can be lost by the fast particle in a single inelastic collision (equation (2)). As a result, if one were to perform the idealized experiment of passing a large number of completely identical particles through such a 'thin' detector, then the energy-loss resulting from the combined effect of the individual ionizing events would be found to have the characteristic skew probability distribution illustrated in figure 2.

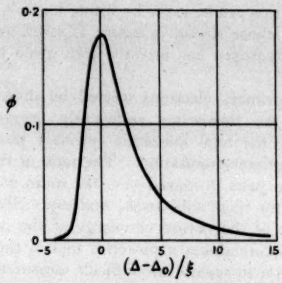


Figure 2. Probability ϕ of an energy-loss Δ as a function of the most probable energy-loss Δ_0 and the parameter ξ defined by equation (12). The area under the curve is normalized to unity. (Laudau 1944).

It is a simple matter to give a qualitative explanation for this shape: because of the $d\epsilon/\epsilon^2$ law of energy transfer mentioned above the great majority of the many ionizing collisions suffered by the particle while traversing the counter will be those in which only a small amount of energy (less than a quantity which we shall denote by ξ) is transferred to the struck electron; such low-energy encounters give rise to the almost gaussian peak. On the other hand, the high-energy tail is predominantly due to the less frequent collisions in which a large amount of energy ($>\xi$) happens to be given to the electron. If we define ξ by equation (12), then on the average just one ionizing event in which an amount of energy greater than ξ is transferred will occur during the passage of a particle through an absorber. If there were only low-energy collisions, one might expect from the statistical nature of the ionization process that the peak would become relatively narrower as the total number of ionizing events is increased—that is, as the detector is made thicker. But increasing the thickness also increases the probability of the events that contribute to the high-energy tail, so that the width of the resultant distribution is in fact very insensitive to the total energy transferred (see § 4.1).

Under the two conditions (*a*) that the energy-loss of a particle in passing

through an absorber of thickness x is small compared with its initial energy, and (*b*) that ξ is large compared with the average ionization potential $I = I_0 Z$ of the electrons in the absorber, Landau predicted the energy-loss distribution shown in figure 2. The two parameters of this distribution Δ_0 (the most probable energy-loss) and ξ are given by the following formulae:

$$\Delta_0 = \xi \left[\ln \frac{3000\,\beta^2 \xi}{Z^2(1-\beta^2)} + 1 - \beta^2 \right] \text{ ev} \qquad \ldots\ldots(11)$$

$$\xi = \frac{2\pi N e^4 \rho}{mv^2} \frac{Z}{A} x = \frac{1{\cdot}54 \times 10^5}{\beta^2} \rho x \frac{Z}{A} \text{ ev} \qquad \ldots\ldots(12)$$

where the symbols have the same meanings as before. Equation (11) (which assumes that ξ is in ev and that $I/Z = I_0 = 13{\cdot}5$ ev) differs somewhat from the expression for the average energy-loss (equation (6)); in particular, Δ_0 is not a linear function of x, owing to the ln ξ term. It will be seen from these expressions and from figure 2 that the relative full width at half-height ($= 4{\cdot}0\xi/\Delta_0$) is dependent only on the logarithm of the absorber thickness. Although derived by completely different mathematical methods, Williams' (1929) results are very nearly identical with those of Landau. The theory has been extended to the case of thick absorbers, in which the particle loses a large fraction of its initial energy, by Symon (1948).

These fluctuations in the rate of energy-loss by ionization, and particularly the 'high-energy tail' of the Landau distribution, make the experimental determination of the average ionization per unit path length a difficult matter. The average is very sensitive to the relatively infrequent high-energy collisions, which are difficult to take into account satisfactorily as they tend to saturate the detector or its amplifier. High-energy recoil electrons from such collisions may quite easily have ranges large compared with the size of the detector—a condition which is itself sufficient to complicate any comparison with theory (Gardner 1954). Another difficulty arises in cosmic-ray experiments: in a few per cent of traversals high energy mesons will produce knock-on electrons from the walls of a counter, so that the ionization liberated in the gas is the sum of the ionization caused by the meson and by the 'knock-on' (which may be a heavily ionizing delta-ray). Then again, in work with cosmic-ray mesons the particles may arrive at the same instant as a shower, so that if the shower is not detected by guard counters (Kupperian and Palmatier 1953) or by a cloud chamber (Becker *et al.* 1952) a spuriously large burst of ionization would be recorded. All of these mechanisms may distort the shape of the 'tail', but will have a very much smaller effect on the gaussian region of the peak, so that it is usually preferable to determine the mode of the ionization distribution rather than the mean.

The Landau theory is in agreement with experiment regarding the value of Δ_0, but under some conditions gives only a very approximate value of the width of the distribution (see § 4.1). Fano (1953) and Hines (1955) have recently pointed out the existence of an approximation in the transport equation on which the Landau theory is based. From the physical conditions to be considered when deriving the equation the upper limit of integration over ϵ, the energy-loss in an individual collision should be the smaller of two quantities: (*a*) the maximum transferable energy T or (*b*) the energy-loss which the particle has suffered before reaching the point in question. In practice (*a*) applies for slow protons and

(*b*) for fast electrons and gaseous detectors. Hines has shown that the width of the distribution is quite sensitive to the upper limit chosen for ϵ. Thus it is clear that large errors may be involved in integrating, as Landau does, from zero to infinity; and also that Landau's expression of results for all particles in terms of a single universal function is an oversimplification. We shall return to a discussion of these fluctuations in energy-loss in § 4.1.

§ 3. The Density Effect

We have so far assumed that the atoms in the medium traversed by a fast charged particle behave as though they were isolated, and were uninfluenced by the presence of their near neighbours. Equation (6) shows that on this assumption the mean energy-loss will increase logarithmically with momentum for values of β greater than 0·97, half of this rise being due to the increase of the maximum transferable energy T and half to the increasing radius of action of the electric field of the particle under the effect of the Lorentz contraction.

It was pointed out by Swann (1938) that the screening effect of polarization in condensed media will reduce the ionization loss due to distant collisions, so that the relativistic increase in energy-loss is less than that given by the simple formula (6); this is illustrated in figures 3 and 4. The average energy-loss shown in

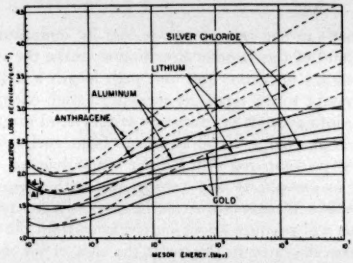


Figure 3. Calculated average ionization loss of μ-mesons in various condensed materials. The broken curves give the values of $(1/\rho)\,(dE/dx)$ which would be obtained without the density effect (Sternheimer 1952).
(*Reproduced by permission of the Editor, Physical Review*)

these figures depends to a considerable extent upon the close collisions which are the cause of the high-energy tail of the Landau distribution (figure 2), and which are hardly affected by polarization. The most probable energy-loss is, on the other hand, mainly determined by the distant collisions, and the effect of polarization is to eliminate altogether the relativistic rise beyond a value depending on the properties of the medium (figure 10).

In the first theoretical treatment of the density (i.e. polarization) effect (Fermi 1939, 1940), the electrons of the medium were regarded as classical oscillators set in motion by the passage of the charged particle, the energy given to the electrons being equal to that lost by the particle. Since polarization effects are important only at distances large compared with the interatomic spacing, Fermi limited his investigation to losses occurring at relatively large distances from the track of the

particle. This limitation allowed him to consider the medium as a continuum, and the energy-loss as a classical process—a simplification which is not valid for close collisions, for which a quantum-mechanical treatment is essential. In this classical treatment the energy-loss occurring at distances greater than some distance b was equated to the flux of the Poynting vector over the surface of a cylinder of radius b and having the track of the particle as axis. The effect of

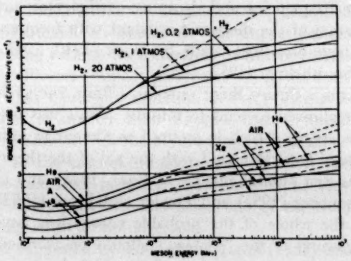


Figure 4. Calculated average ionization loss of μ-mesons in various gases. The broken curves give the values of $(1/\rho)\ (dE/dx)$ which would be obtained without the density effect (Sternheimer 1952).

(*Reproduced by permission of the Editor, Physical Review*)

close collisions was allowed for by joining this calculation on to a Bethe–Bloch calculation at some suitable small value of b, where the effect of polarization was vanishingly small. The overall reduction of energy-loss due to polarization was then represented by the difference between the energy-losses calculated (i) for finite polarization as above, and (ii) for negligible polarization, using the normal Bethe–Bloch methods. A further simplification was the assignment of only one characteristic frequency ν to the electrons of the medium.

This treatment yielded the following results: the reduction in the energy-loss per unit path length resulting from polarization is

$$U = \frac{2\pi ne^4}{mv^2} \ln \eta \quad \text{for} \quad v < c/\sqrt{\eta} \qquad \ldots\ldots(13)$$

and

$$V = \frac{2\pi ne^4}{mv^2}\left[\ln \frac{\eta - 1}{1 - \beta^2} + \frac{1 - \eta\beta^2}{\eta - 1}\right] \text{ for } v > c/\sqrt{\eta} \ \ldots\ldots(14)$$

where $n = N\rho Z/A$ is the number of electrons per unit volume, and $\eta = 1 + 4\pi ne^2/mv^2$ is the dielectric constant of the medium for low frequencies.

The quantity U is very small, particularly for gases, so that in the low-energy region Fermi's theory does not give results appreciably different from the Bethe–Bloch theory. On the other hand, at high energies, when the particle velocity exceeds the velocity of light in the medium, the correction (14) has to be subtracted from formula (6) for the average or from expression (11) for the most probable energy-loss. In the case of the latter the factors $1 - \beta^2$ in the logarithms cancel, so that the probable energy-loss tends towards a constant value, the so-called ' Fermi plateau ', independent of the particle velocity.

The Fermi theory was later extended by Halpern and Hall (1940, 1948) to cover the case of several dispersion frequencies. For the two limiting cases of very high ($v \simeq c$) and very low energies the single-frequency and the multi-frequency theories show no significant difference, but the latter gives a much more gradual transition between the logarithmic portion of the Bethe–Bloch energy-loss curve and the limiting ' Fermi-plateau ' value. Moreover, the exact shape of the curve in this transition region depends critically on the assumptions regarding oscillator frequencies and damping coefficients; or, in more physical terms, on the variation of the dielectric constant with frequency. Extensions of the Fermi theory have also been made by Wick (1940, 1941, 1943), Schönberg (1950, 1952), and Sternheimer (1952, 1953).

According to Fermi's theory there remains a finite energy-loss even when the limiting radius b is allowed to tend to infinity. This was interpreted by Fermi as that part of the energy which is emitted as Cerenkov radiation, which then forms part of the total loss calculated with the aid of the theory, and is not additional to it. It was first pointed out by A. Bohr (1948), and later by Messel and Ritson (1950), Schönberg (1951) and Fowler and Jones (1953) that on the basis of Fermi's theory the whole of the probable energy-loss beyond the minimum and up to the beginning of the ' plateau ' should be ascribed to this Cerenkov radiation. In this energy region, where $\beta < \eta^{-1/2}$ the energy-loss found from Fermi's calculations when b tends to infinity is

$$-\left(\frac{dE}{dx}\right)_{\infty} = \frac{2\pi ne^4}{mv^2}\left[-\beta^2 - \ln(1-\beta^2)\right] \qquad \ldots\ldots(15)$$

so that the probable energy-loss excluding Cerenkov radiation, which is found by subtracting both (13) and (15) from (11), reduces to a constant value, equal for unit thickness to

$$-\left(\frac{dE}{dx}\right)_{\text{ion}} = \frac{2\pi ne^4}{mv^2}\left[\ln\left(\frac{3000}{Z^2\eta}\frac{2\pi ne^4}{mc^2}\right) + 1\right]. \qquad \ldots\ldots(16)$$

This equation does not contain β, and therefore implies that there should be no relativistic increase in energy-loss, except for that portion given off as Cerenkov radiation. As most of this radiation would not be absorbed in the vicinity of the track, one would expect on the basis of Fermi's original theory to find no relativistic increase in probable *ionization*.

The first results obtained with photographic plates appeared to show no increase of grain density at high energies. They were supported by some measurements of Bowen and Roser (1951) with an anthracene scintillator, though it is also true that a detectable relativistic increase in ionization in other media had been reported earlier. The more recent experimental results of (among others) Ghosh, Jones and Wilson (1952, 1954), and of Pickup and Voyvodic (1950, 1951) have established the existence of a relativistic increase of ionization, which is certainly connected with events occurring close to the track of the particle. In view of this discrepancy between theory and experiment the distribution of energy-loss between ionization, excitation and Cerenkov radiation has therefore been reinvestigated in a number of recent theoretical papers.

Hüybrechts and Schönberg (1952) begin by pointing out that the procedure adopted by Fermi for overcoming the difficulty of treating close collisions may give misleading results, since the frequency bands in which coherence of the

emitted Cerenkov radiation is possible (and in which therefore a large part of the emitted Cerenkov energy will be located) will be modified by interactions between adjacent atoms. They overcome the difficulty by dividing their calculation into two parts: up to a radius $R = (mc^2/4\pi ne^2)^{1/2} = 5{\cdot}3 \times 10^5\, n^{-1/2}$ cm, polarization effects are relatively unimportant, and the time of collision is short; resonance effects therefore tend to be blurred by the uncertainty principle so that the usual treatment of dispersion (Heitler 1947), is inapplicable. In this region, therefore, they apply the Bethe–Bloch Møller–Williams theory, corrected for the effect of polarization occurring outside a distance R. At impact parameters greater than R collision times are long enough for a classical treatment to be applicable. They conclude that close collisions (impact parameter $< R$) lead mainly to ionization and excitation, and that Cerenkov radiation is the principal result of the more distant collisions, and is smaller than Fermi's theory would indicate. Their results are in good agreement with the observations of Ghosh *et al.* (1954, fig. 10). On the other hand, in the case of photographic emulsion their theory predicts a smaller relativistic rise than is found experimentally ; they suggest that the discrepancy may be due to the fact that small energy transfers are probably not effective in developing silver bromide grains.

It has since been suggested by Fowler and Jones (1953) that it is unnecessary to adopt the somewhat artificial device of treating collisions with impact parameters less than R differently from the more distant collisions. They point out that the main interaction is in any case electrostatic, so that for distances greater than atomic dimensions ($\sim 10^{-8}$ cm) the medium should be treated as a homogeneous dielectric. The difficulty regarding resonance effects at short collision times disappears when the terms due to the continuous spectrum of states of the atomic electrons are included. If this is done it is found that the dielectric constant is complex over the whole range of frequencies corresponding to the continuous spectrum, and that this gives rise to a heavily damped Cerenkov radiation, which is entirely absorbed within a distance of the order of 10^{-4} cm from the track of the particle. As far as macroscopic effects are concerned, therefore, their results are similar to those of Hüybrechts and Schönberg's ' two-region ' theory.

In the mechanism considered by Budini (1953) a similar distinction is made between ' primary ' excitation and ionization, and ' secondary ' effects caused by the re-absorption of Cerenkov radiation. He shows that the whole of the increase of energy-loss beyond the minimum should appear as Cerenkov radiation only in the special case of a perfectly transparent medium, i.e. when the width of the spectroscopic lines is negligible, but that this is a physically impossible condition. His theory, like that of Hüybrechts and Schönberg, is able to account satisfactorily for the large increase in ionization above the minimum which is observed in gases. The exact form of the transition between the logarithmic increase region and the Fermi ' plateau ' and the details of the emission of Cerenkov radiation are both found to depend on the ratio

$$\frac{\text{breadth of the spectroscopic lines}}{\text{density of the medium}};$$

this quantity plays an important role in his calculations (figures 10, 13). He points out that one would expect a correlation between the breadth of the spectroscopic lines and the structure of the Cerenkov bands which should be capable of experimental test. His work leads to the further conclusion that for dense media

with narrow spectroscopic lines one should be able to observe a relativistic increase of the Cerenkov radiation simultaneously with a relativistic increase of ionization. Such an effect has been reported by Bassi *et al.* (1952).

Considerable contributions to the multi-frequency dispersion treatment of the density effect have been made by Sternheimer, particularly as regards the detailed correlation of experimental results with this theory. Figures 3 and 4 reproduced from his 1952 paper are typical of the results obtained for gases. The dependence on pressure shown there may be expressed in words: the density-effect correction at one atmosphere for a particle of momentum p is the same as that at P atmospheres for a particle of momentum $p/\sqrt{P}$. (The ordinates in figures 3 and 4 represent the *average* ionization loss, which, as mentioned above, continues to increase without limit in the extreme relativistic region, even though the most probable ionization reaches a limiting value.)

More recently (1953 a, b) Sternheimer has made a detailed investigation into the relative importance of ionization and Cerenkov radiation in silver bromide grains, and has concluded that the Cerenkov loss is only one-fiftieth of the ionization loss. For gases it is somewhat more important, particularly in the case of helium; his calculations for the case of oxygen are compared in figure 10 with the experiment of Ghosh *et al.* and also with Budini's theoretical results. In his most recent paper (1954), he points out that a correction ought to be applied to the accepted value of the average ionization potential as determined from the stopping power of a particular material for fast charged particles. This correction is in principle necessary because the stopping power will have been modified by the density effect; however, he is able to show that the correction is a very small one.

The Fermi theory has also been applied to the case of the photographic plate by Allen (1954). By inserting known values for the photoelectric absorption coefficient of silver bromide into the theory, he shows that the Cerenkov radiation contributes very little to the energy-loss—a conclusion which is in agreement with that of Sternheimer.

The present state of agreement between experiment and the various theoretical treatments is typified by figure 10: while the general agreement is good, the results are not yet sufficiently precise to allow a final choice between the various theoretical alternatives to be made. In order to elucidate the fine details of behaviour it will be necessary (as Ghosh *et al.* point out) to abandon cosmic-ray experiments with their low counting-rates, and to concentrate on utilizing the intense fluxes of high-energy electrons that are now available.

§ 4. Results obtained with Gaseous Detectors

In reviewing the experimental work, it is convenient to begin with effects in gases, since it is there that the logarithmic rise is most easily observed, free from the complications of the density effect. Moreover, it is logically desirable to prove the existence of this rise before going on to argue that a much smaller effect in condensed media is evidence for the density effect.

4.1. *Proportional Counters and Pressure Ionization Chambers*

An early attempt at measuring the specific ionization of cosmic-ray particles with two proportional counters was made by Weisz and Swann (1946) but it appears to have been abandoned when it was realized how serious was the effect

of the fluctuations mentioned in § 2.3. There appears to be no record of any further work for several years. Meanwhile, however, the proportional counter was coming into general use as a precision instrument, so that by 1950 attention was once more focused on the possibilities of this method. Two factors contributed to this development: the greatly enhanced stability of modern amplifiers and high tension supplies, and the increased use of x-rays, particularly K-capture x-rays from artificially activated radio-isotopes, for energy calibration (e.g. Rothwell and West 1950, Insch 1950, Cranshaw 1951). Some idea of the versatility of the present-day instrument may be gained from a recent review article by West (1953b).

The experiments with relativistic μ-mesons that have been conducted up to the present time all follow the same general pattern: one or more proportional counters are placed in the beam of a Geiger-counter telescope, and measure the ionization produced, while the energy of each particle passing through the system is determined either by its residual range, or by deflection in a magnetic field. A typical arrangement is shown in figure 5. It is necessary to assume a know-

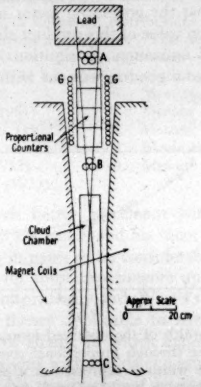


Figure 5. Typical experimental arrangement for the determination of the probable ionization of cosmic-ray mesons. Four proportional counters are used. Counters A B C define the beam. The proportional counters are protected from the effect of showers by the guard counters G. The momentum of the particle is measured with the cloud chamber and magnetic field (Price *et al.* 1953).
(*Reproduced from the Proceedings of the Physical Society*)

ledge of the nature of the particle, but at sea-level there is no difficulty in eliminating virtually all particles other than μ-mesons. Although some refinements of counter technique are possible, in the main it is the quality of the momentum measurement that determines the value of the results: range measurements (Chanson *et al.* 1952, Kupperian and Palmatier 1953, Snodgrass 1953) are useful with μ-mesons only up to momenta around $1{\cdot}5 \times 10^9$ ev/c, which is nowadays too low to be of real interest; large cloud chambers in magnetic fields are usable with some difficulty up to 5×10^9 ev/c (Becker *et al.* 1952, Price *et al.* 1953).

Higher energies require the use of counter-hodoscope arrays in which particles are deflected in intense magnetic fields, and the deflection is measured by counter trays several metres apart (Parry *et al.* 1953, Yeliseyev *et al.* 1953, Eyeions *et al.* 1955, Kharitonov *et al.* 1952, 1953). Three such cosmic-ray spectrometers, at Melbourne (Caro *et al.* 1950, 1951), Manchester (Hyams *et al.* 1950) and Alagez, U.S.S.R. (Alikhanyan 1951) have been used recently for studying the ionization of extreme-relativistic mesons. The Manchester instrument has an additional refinement, in the form of small cloud chambers, which enables the deflection of very high energy particles to be measured with greatly increased precision, and which raises the maximum observable momentum to 10^{12} ev/c.

The accuracy with which the most probable ionization of a homogeneous group of particles can be identified is proportional to the square root of the number of observations divided by the width of the distribution. Several authors have used more than one counter in an effort to increase the number of observations; four were employed by Price *et al.*, but this seems to be about the practical limit. It is therefore worth enquiring to what extent one can influence the width of the ionization distribution by altering the counter dimensions and filling. It has been shown by West (1953 a) that the governing factor is the product of gas pressure and track length: for a given value of this product all counter fillings give approximately the same width of ionization distribution (figure 6). As may be seen from figure 7 his results show good agreement with those obtained in a number of experiments with mesons.

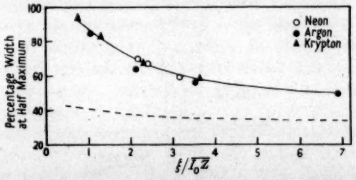


Figure 6. Values of the full width of the observed ionization distribution for minimally ionizing electrons passing through proportional counters with various fillings, as a function of ξ/I_0Z. The width is expressed as a percentage of the most probable ionization. A small correction has been applied for broadening introduced by the counter multiplication process (West 1953 a).
(*Reproduced from the Proceedings of the Physical Society*)

Since the observed widths approach asymptotically to Landau's calculated value at high pressures it is instructive to rewrite one of the conditions of validity of his calculation ($\xi \gg I_0Z$). This may be very easily shown to be equivalent to the following condition, true for all gases:

$$\Pi = (\text{pressure in atmospheres}) \times (\text{depth in cm}) \gg 2. \qquad \ldots\ldots(17)$$

Experimentally it is found that the observed width begins to approximate to Landau's value when $\Pi \sim 70$ (Bradley 1955). In view of condition (17) the inability of Landau's theory to explain the results for smaller values of Π is not at first sight surprising. However, it has recently been shown by Moyal (private communication to be published later) that the range of validity of Landau's theory

with regard to this condition is greater than was formerly thought. In view of this conclusion, the broad distributions observed experimentally strongly support the suggestion of Hines and Fano (mentioned earlier) that there is a defect in the Landau theory. Although the improvements Hines has shown to be necessary have not yet been extensively tested against experiment, the results so far obtained are encouraging. For instance, whereas Landau's distribution is too narrow for electrons, and much too broad for slow protons (Reynolds *et al* 1953), Hines'

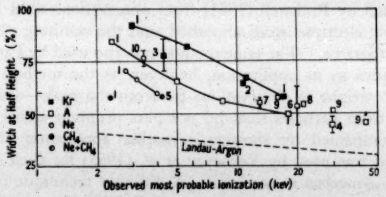


Figure 7. Comparison of widths of ionization distributions obtained by various authors for minimally ionizing electrons and mesons. The widths have been corrected for the broadening introduced by the multiplication processes. Bradley's results extend to an energy-loss of 144 kev when the width has fallen to 37%.

1.	Electrons	West (1952)	6.	Mesons	Cranshaw (1951)
2.	Electrons	Rothwell (1951)	7.	Mesons	Parry *et al.* (1953)
3.	Mesons	Becker *et al.* (1952)	8.	Mesons	Snodgrass (1953)
4.	Mesons	Kupperian and Palmatier (1952)	9.	Electrons	Bradley (1955)
5.	Mesons	Price *et al.* (1953)	10.	Mesons	Eyeions *et al.* (1955)

theory appears to be in much better agreement with experiment. For slow protons his results are more or less exact, and his theory also requires the energy-loss distributions for electrons in gases to be considerably broader than Landau's.

The counter fillings used for most experiments up to the present have been chosen for convenience, and not primarily with West's results in mind. The lack, until recently, of a rigorous theory applicable to such counter fillings has led to a somewhat illogical state of affairs, in which experimentalists compare their observed most probable ionizations with Landau's caculated values of energy-loss even for conditions in which his theory cannot be expected to be strictly valid. Some justification for this procedure may be found from West's measurements on electrons, which show that the ratio of experimental to calculated most probable ionizations is not sensitive to gas pressure; on the other hand he finds that this ratio is appreciably less than unity, with a value of $0{\cdot}89 \pm 0{\cdot}03$.

In the absence of a comprehensive theory of ionization fluctuations several groups had planned to take the obvious next step of making their counters more nearly satisfy condition (17), by using higher pressures and longer path lengths (e.g. Palmatier *et al.* 1954). High-pressure chambers can be operated quite satisfactorily as electron-collection chambers. On the other hand, the high density of the gas itself imposes a limitation on investigations of the relativistic increase of ionization, due to the onset of the density effect at lower particle energies; this may be seen from figure 4.

For the reasons set out in § 2.3 most experiments attempt to measure the change of most probable ionization with momentum from the region of the minimum

upwards. This change, before the plateau is reached, is about 50%, which is only of the same order as the width of the observed ionization distributions. This in itself would cause no difficulty if sufficient observations were available; but this is all too rarely the case. Most experimenters are therefore left with the problem of determining the most probable ionization from a set of irregular histograms representing rather few observations. It is then clearly essential to use an objective method of analysis. This was recognized by Becker *et al.*, who used an application of the method of maximum likelihood (Behrens 1951): the ionization distributions obtained by Rothwell (1951) from observations on a large number of minimally ionizing electrons were smoothed, and the resulting curve fitted to the peaks of their histograms. The same method was also used by Parry *et al.* There are some difficulties in its application, however, as the method of computation gives appreciable weight to observations far from the peak—observations which may have come from particles that did not pass properly across the counter, or which were accompanied by showers. Another application of the maximum likelihood method was used by Yeliseyev *et al.* (1953) for their measurement of the ionization of μ-mesons in methane. A different technique (using linear unbiased minimum variance estimators) was applied by Hammersley and Morton (1954) to the analysis of the results of Eyeions *et al.*; this method gives negligible weight to the points far from the peak, and can be reduced to the mechanical application of a ' computing recipe '.

All these methods use the whole of the experimental data, and are therefore to be preferred to the ingenious, but somewhat rough-and-ready way of sharpening up the experimental ionization distribution proposed by Landau, and used by Nikitin (1948). This takes advantage of the fact that the calculated ' Landau ' distribution shows a very sharp cut-off on the low-energy side of the peak, so that if two observations on each particle are available, and the lower one only is selected; then these selected observations would be expected to have a smaller variance, and a mode that is displaced very little from the mode of the unselected observations.

With one exception (since withdrawn) all the published experiments show the expected relativistic increase; but up to the time of writing the published investigations of the high-energy plateau region have been limited to the two sets of results shown in figures 8 and 9. Although a number of other experiments show an apparent falling-off in the rate of rise of the ionization curve as the limit of momentum resolution is reached, it is not always certain that this is a genuine effect: for instance, the combination of a power-law meson spectrum and a gaussian error in the determination, from cloud chamber photographs, of the curvature of a track in a magnetic field, can give just such a spurious result. For the existence of the plateau to be considered as established, the maximum observable momentum should be well above the value at which the plateau is reached. Other possible sources of error, both giving the same spuriously reduced rate of rise, are (i) saturation of the proportional counter due to the use of too high a gas amplification and (ii) variation of counter sensitivity across the sensitive volume, owing to recombination. Experiments which claim to establish a rate of rise that is smaller than expected theoretically lose a great deal of conviction if they do not include a set of x-ray calibrations designed to show that the counter is linear in response, and has a constant amplification factor throughout its sensitive volume. The results shown in figures 8 and 9 satisfy these criteria.

4.2. *Cloud Chambers*

The use of cloud chambers to measure ionization is one of their best-known applications, especially in the field of cosmic-ray physics. For many purposes an approximate measurement—correct to a factor of about two—is sufficient, and this can be achieved with very little difficulty. However, the detection of the relativistic rise involves working to accuracies of within a few per cent, and this proves to be a very different matter.

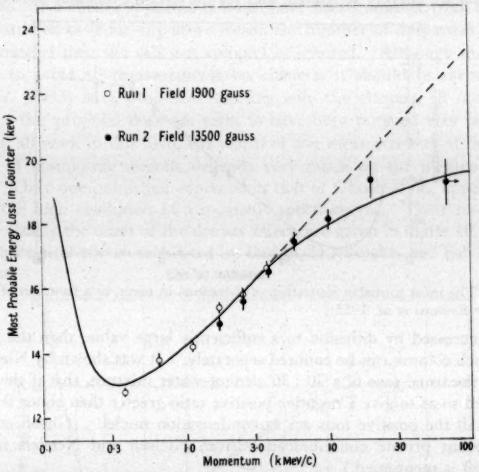


Figure 8. The most probable ionization of mesons in argon as a function of momentum (after Parry *et al.* 1953).
(*Reproduced from the Proceedings of the Physical Society*)

Before mentioning the precautions that have to be taken it is desirable to distinguish between the two types of measurement that can be made. If the expansion takes place immediately the ionizing particle has passed, so that the ions have not had time to diffuse, then either a single drop of liquid forms on the group of secondary ions produced by the slowing-down of each primary recoil electron, or there is a group of drops so close together that they cannot be resolved. In either case it is possible to count with fair accuracy the total number of ionizing collisions suffered by the incident particle. The number of such events per centimetre at S.T.P. is identical with the primary specific ionization discussed in §§ 2.2 and 4.3. Since the measurements with low pressure counters and cloud chambers are thus complementary, the results from both types of experiment are summarized together in the following section (§ 4.3).

If, on the other hand, the expansion of the chamber is delayed by a few hundredths of a second, then the ions will have had time to become separated by diffusion, and each ion will form the nucleus of a separate drop. Over most

of the track the individual drops can then be resolved, and the most probable ionization can be found by a statistical analysis of the drop density.

Whichever type of measurement is being undertaken, it is quite essential to ensure that throughout the period of the experiment the fraction of ions acting as condensation nuclei is constant. This is most easily done in measurements of the probable ionization by taking advantage of the fact that drops condense more easily on positive than on negative ions. If an electrostatic clearing field is applied before a slightly delayed expansion, so that (*a*) the positive and negative ion columns are slightly separated, and (*b*) the average distance between the ions

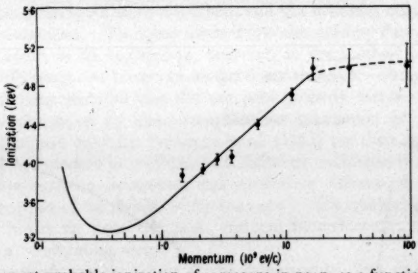


Figure 9. The most probable ionization of μ-mesons in neon, as a function of momentum (*after* Eyeions *et al.* 1955).

has been increased by diffusion to a sufficiently large value, then the number of drops in each column can be counted separately. It was shown by Nielsen (1941, 1942), for the usual case of a 70 : 30 alcohol–water mixture, that if the expansion is controlled so as to give a negative/positive ratio greater than about 0·2 : 1, then effectively all the positive ions act as condensation nuclei. (Ghosh *et al.* (1954) quote a recent private communication from Nielsen and Needels in which a figure of 0·3 is mentioned.)

This method of obtaining a constant condensation efficiency depends on splitting the track, and allowing time for diffusion. If this is undesirable for other reasons—if, for instance, the same track is being used for momentum measurement, or if the primary ionization is the quantity of interest—one has to resort to different methods. Carter and Whittemore (1952), in their measurements of the probable ionization of mesons, operated a cloud chamber under very carefully controlled, constant, expansion conditions, relying on their operating technique to give them sufficiently constant condensation efficiency ($\sim 75\%$).

Corrections to the drop count have to be applied for the general slight background of condensation, for drops associated with the tracks of other particles, and for drops which escape detection because they cannot be resolved from others. The measurements are also complicated by the fluctuation phenomena discussed in connection with proportional counters: the relatively infrequent high-energy collisions are observed as a dense cluster of overlapping drop images, the detector (the photographic plate) being effectively saturated. As there is no hope of estimating the number of drops in such a cluster with any accuracy, it is impossible to make a satisfactory determination of the average ionization. Once again, therefore, it is preferable to measure the probable ionization.

The effect of the clusters has to be eliminated in some way: it is usual to reject clusters containing more than about 30 drops (which corresponds to setting an upper limit to the energy transfers considered of about 1 kev). Although equation (3) shows that the observed value of the ionization is not sensitive to the exact choice of this limiting value, it is nevertheless highly desirable when making this selection to be guided by some suitably objective rule. The work of Ghosh *et al.* (1952, 1954) is in this respect admirable. They ruled a series of parallel lines on their photographic plates, dividing the tracks into a number of equal-length cells. If a cluster appeared in a cell (and also if another track crossed the one being studied or if for any other reason the number of drops was difficult to count accurately) then the cell was completely ignored. Although most authors endeavour to avoid all measurements on clusters, it should be mentioned that Cohn *et al.* (1953) have proposed counting *only* the clusters, in order to save labour; but this proposal does not seem to have been pursued very far.

As with all work in this field, the worth of any measurements of the probable ionization of cosmic-ray mesons depends very much on the momentum determination. Only one published experiment, that of Ghosh *et al.*, appears to have exploited the high resolution of a magnetic spectrometer. Their results, which show very clearly the onset of the density effect, are given in figure 10, which also includes theoretical curves calculated by Budini, Hüybrechts and Schönberg, and Sternheimer.

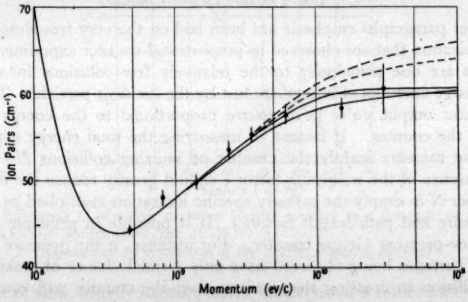


Figure 10. Average specific ionization of μ-mesons in oxygen at S.T.P. arising from encounters of energy transfer less than 1 kev.

The experimental points are the data of Ghosh, Jones and Wilson (1952, 1954). The full curves are for two extreme assumptions made by Budini (1953) regarding the width of spectral lines; they are normalized to the experimental value near the minimum of ionization. The lower curve coincides almost exactly with one calculated by Hüybrechts and Schönberg (1952). The broken curves are from a calculation by Sternheimer (1953); the upper curve gives the rate of energy-loss, and the lower the energy deposited close to the track after subtraction of the estimated Cerenkov loss. These curves show the extent to which the available measurements fail to distinguish between details of theoretical treatment.

The other work on mesons that has been published suffers from relatively poor momentum determinations: Carter and Whittemore (1952) were unable to extend their measurements in a helium chamber to momenta greater than

$1{\cdot}5 \times 10^9$ ev/c (which is hardly high enough to make the relativistic rise easily detectable), and based their claim to have verified the effect on the observation that the ionization of four high-energy electrons was appreciably greater than the minimum. Hazen (1944, 1945) obtained somewhat inconclusive results for mesons, but verified the rise for electrons. Similarly Sen Gupta's (1940) results showed less than the expected rise for mesons, although for electrons he was able to verify the theory up to energies of the order of $1000mc^2$; unfortunately, the accuracy of his two highest points was not good enough to establish the existence of a Fermi plateau. Earlier, the measurements of Corson and Brode (1938) had provided convincing evidence for the rise; they used cosmic-ray electrons of energies up to 30 MeV, and a nitrogen-filled cloud chamber. More recently, Hayward (1947) has shown that very high energy electrons have an ionization in helium consistent with a plateau value of the order of $1{\cdot}4\ I_{\min}$. The careful work of Frost and Nielsen (1953), with argon, air and helium, was confined to electron energies less than 1·7 MeV. Beekman's (1949) measurements of the specific ionization of electrons in a mixture of air and ethyl alcohol also did not extend beyond the minimum of ionization. Thus there appears, rather surprisingly, to be no single published experiment with electrons in which the whole range of interest is covered with the same thoroughness that characterizes the work of Ghosh *et al.* on mesons.

4.3. *Low Pressure Geiger Counters*

In earlier paragraphs emphasis has been laid on the very troublesome fluctuations of ionization that are observed in proportional counter experiments. These fluctuations are due principally to the relatively few collisions in which comparatively large amounts of energy are lost by the incident particle. The proportional counter output pulse is of course proportional to the energy $\int W(\epsilon)\ \epsilon\ d\epsilon$ released in the counter. If instead of measuring the total energy expended, we were able to measure simply the number of ionizing collisions $N = \int W(\epsilon)\ d\epsilon$, then the absence of the weighting factor ϵ should greatly reduce the fluctuations. (The number N is simply the primary specific ionization multiplied by the appropriate pressure and path-length factors.) It is possible in principle to measure N with a low-pressure Geiger counter. For instance, if the pressure is adjusted so that the particles being observed have only a small chance of making a single ionizing collision in crossing the counter, then the counter will count just the number of ionizing events. Somewhat higher pressures can be used if it is assumed that the distribution of ion pairs per traversal follows Poisson's law: if the average number of ionizing collisions per traversal is q, then the probability of no pair being produced is e^{-q}. The efficiency C of the counter is therefore $1 - e^{-q}$, provided (*a*) the counter is sensitive to the presence of a single ion pair and (*b*) there are no ' knock-on ' electrons emerging into the counter gas from the walls.

The first of these conditions is satisfied sufficiently well in practice if gases with low attachment coefficients are used, and if the counters are operated on their plateaux. The accuracy with which the second is met is usually investigated experimentally by measuring the counter efficiency as a function of the gas pressure. If there are no electrons knocked-on from the walls, then plotting log $(1 - C)$ against pressure should give a line passing through the origin; if this line is also

straight, then the use of Poisson's law is certainly justified. In practice extremely good straight lines passing through the origin are obtained (see, for instance, figure 11, reproduced from a paper by McClure 1953); this is usually taken as indicating that the walls make no appreciable contribution. However, this converse argument is not strictly true; for instance, if on the average k 'fast' knock-on electrons—i.e. of energy greater than about 1 MeV—are produced in the walls of a counter by a fast meson, then as each produces about as many ion-pairs per centimetre (on the average) as the primary particle, the effect will be to increase the apparent value of q by a factor $1+k$. Any error of this kind is not detectable by the method of extrapolating to zero pressure as in figure 11, *since it will vary with pressure in exactly the same way as the wanted effect.*

Fortunately, the value of k is small: the observations of Hazen (1943), Brown *et al.* (1949, 1950), Walker (1953), Wolfendale (1953) and Lovati *et al.* (1954), and the calculations of Hereford (1949) all show that k is less than about 0·08–0·10 even for quite thick-walled counters; but the point is worth bearing in mind when the highest accuracy is being sought.

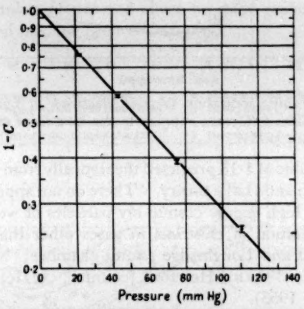


Figure 11. Efficiency C of a low pressure Geiger counter as a function of gas pressure for 0·855 MeV β-rays (McClure 1953).
(*Reproduced by permission of the Editor, Physical Review*)

The case of slow, heavily ionizing knock-on electrons is rather different as they would be detected with about 100% efficiency by the counter, for practically any pressure, and so would tend to prevent the line in figure 11 from passing through the origin. The fact that it does so shows that such slow knock-ons very rarely escape from the counter walls. This is independently confirmed by Wolfendale (private communication) who finds that only $\frac{1}{2}$% of ionizing particles emerging from a lead block are accompanied by knock-on electrons of energies between 30 kev and 1 MeV.

In figure 12 are shown the results of four measurements of the primary specific ionization of electrons in hydrogen, including the cloud chamber observations of Williams and Terroux (1929). The experimental values of the different authors are considerably scattered, and, except for Hereford's (1948) results, provide no evidence for the relativistic increase of primary ionization. There is rather better agreement regarding the value of the primary ionization in hydrogen of the mixed

sea-level cosmic radiation: Danforth and Ramsey (1936), Cosyns (1936) and Hereford all obtain values close to 6·0 ions per cm. This is about 1·2 times McClure's minimum value for electrons, which is roughly what would be expected from the known sea-level spectrum if the relativistic rise exists. Further confirmation is provided by the experiment of Shamos and Hudes (1951), who exposed a hydrogen filled counter to the sea-level cosmic-ray meson spectrum, and later to the spectrum filtered by 140 feet of rock. The experimental ratio (I_{140}/I_0) was 1·17 $\pm$ 0·03,

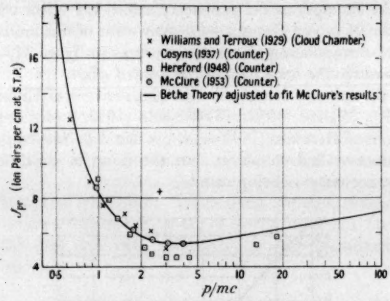


Figure 12. Primary specific ionization (J_{pr}) of electrons in hydrogen, as a function of momentum p. The full curve represents the best fit of the Bethe formula to the results of McClure (1953).

which is close to the value of 1·18 predicted theoretically from Wilson's (1946) meson spectrum and Halpern and Hall's theory. There do not appear to be any published experiments utilizing high energy cosmic-ray particles of well-determined energy.

The primary ionization of electrons in gases other than hydrogen has been studied by Skramstad and Loughridge (cloud chamber; N_2, Ne; 1936), Hazen (cloud chamber; He; 1943 a), Hereford (counter; A, He; 1947) and McClure (counter; He, Ne, A; 1953).

§5. Results obtained with Condensed Detectors

5.1. *Photographic Plates*

The introduction by Berriman in 1948 of photographic plates sensitive to singly charged minimally ionizing particles opened up a new method of investigating ionization in solid media. The first reports concerning the relativistic increase were negative: Occhialini (1949) found no detectable variation in grain count for electron energies from 7·5 to 500 Mev, his work being supported by Corson and Keck (1950) who showed that the grain count for electrons of from 10 to 180 Mev was constant to within $\pm 1\frac{1}{2}\%$. Negative results were also reported by the Bristol group at the 1949 Edinburgh conference (*Nature, Lond.*, 1950, **165,** 54). However, the results of Pickup and Voyvodic (1950), published shortly afterwards, made it seem probable that a small difference of about 10% did exist between the plateau level of ionization for extreme relativistic particles and the minimum value, a dip occurring at electron energies just below those studied by Occhialini, Corson and Keck. This work has since been confirmed by a number of other experimenters, whose collected results are given in figure 13.

In seeking to substantiate such small differences of ionization it is important to choose an objective method of counting the exposed grains. The earliest work was based on an attempt to count every grain; but although the results obtained by the best trained observers were self-consistent to $\pm 2\%$, differences of as much as 20% existed between observers (Jauneau and Hug-Bousser 1952). These differences, which arose principally from the difficulty of counting overlapping grains, introduced an unwelcome subjective element into the work. Morrish (1952) therefore proposed the adoption of the ' blob ' counting convention, in which a single isolated grain, or a group of overlapping grains, are both counted as one ' blob '. This method of counting, which is particularly suitable for grain densities up to 1·5 times the minimum, and on which much of the recent work is based, is found to give results that vary by not more than 0·5% from observer to observer. For the region of particular interest between the minimum and the plateau the blob count is proportional to the grain count (Stiller and Shapiro 1953). Della Corte *et al.* (1953) have pointed out that this linear relationship would be expected to break down for blob densities greater than three or four times the minimum value, and have therefore proposed an alternative procedure for estimating the actual number of exposed silver bromide grains; up to the present, however, their method has not been widely used.

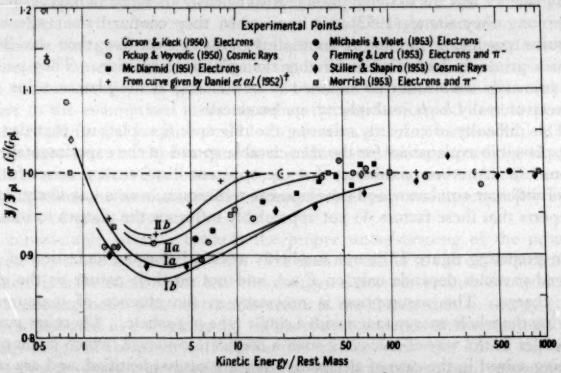


Figure 13. Comparison of calculated values of ionization $\mathfrak{J}$ loss in silver bromide grains and observed grain or blob density (G) in photographic emulsions. All curves are normalized to a plateau value of unity.

† Summarizing their results and those of Voyvodic and of Danysz *et al.* (1952) (cosmic rays).

Theoretical curves. I*a*: Maximum transferable energy $T=10$ kev (Sternheimer 1953); I*b*: $T=2$ kev (Stiller and Shapiro, private communication), based on Sternheimer's theory. Curves I*a* and I*b* take the absorption of Creenkov radiation by silver bromide into account. II*a*, *b*: Calculated on two different assumptions regarding the widths of the absorption lines; $T=5$ kev (Budini 1953).

In order to compare the results of different experiments the points shown in figure 13 have been normalized to give the same plateau values of ionization for high energy electrons. Some such normalization is essential, as the absolute

values of the grain count depend on a large number of factors: temperature of exposure, development conditions, differences between batches of plates, and differences between observers. This particular normalization appears to be justified, as most experiments appear to give no variation in grain or blob counts for particles of $E/\mu c^2$ greater than 40.

In experiments with cosmic rays it is not always completely certain which should be regarded as the standard particles, assumed to be ionizing at the plateau value. If a particle chosen as a standard has been wrongly identified, and is in fact near the minimum of ionization, then a reduced rise will be recorded. The difficulty is that the choice of standard particles depends to some extent on the assumptions made regarding the energy at which the plateau is reached. If it is assumed that the plateau is reached more rapidly than is in fact the case, so that particles of too low an energy are taken as ' standard ', then the result will be that the observed ionization curve will saturate too rapidly. Thus there is the unsatisfactory feature that the result depends on, and tends to agree with, the result expected, even when the latter is incorrect.

It is usual to select ' standard ' particles from those primaries which have produced showers in the plate. However, as Stiller and Shapiro point out in a private communication, the mere fact that a primary produces a shower is not in itself evidence that the ionizing particle is sufficiently energetic to have a ' plateau ' ionization; they found (1952) that even when they confined their attention to primaries producing showers of multiplicity greater than five, then one-third of all such primaries had minimum rather than plateau ionization. They point out that primaries should not be assumed to be ionizing at the plateau value unless showers of really high multiplicity are produced.

This difficulty of correctly selecting the reference (i.e. plateau) particles is the most plausible explanation for the considerable spread of the experimental results (figure 13). Another possibility, that the apparent disagreement arises from the use of different emulsions and development techniques, seems less likely, in view of reports that these factors do not appreciably influence the plateau to minimum ratio.

In preparing figure 13 it was implicitly assumed that the ionization of singly charged particles depends only on $E/\mu c^2$, and not on their nature or the sign of their charge. This assumption is necessary in the absence of measurements covering the whole energy range with a single type of particle. There are practical difficulties in the way of achieving such a complete coverage, which come nearest to being solved in the case of electrons: they are easily identified, and are readily available with a wide range of energies from zero to hundreds of MEV. Unfortunately, in the interesting region close to the ionization minimum multiple scattering becomes very marked. Consequently it becomes difficult to measure the track-length in the emulsion, or to avoid accidentally including ' background ' grains lying close to the track, or to avoid jumping unwittingly from one track to another at points where tracks cross; finally, multiple scattering introduces a high probability that the particle will leave the emulsion after only a comparatively short distance. For these reasons measurements on electrons below about 7 MEV require very great care if they are to be trustworthy, although they have been extended to as low as 2 MEV by McDiarmid (1951).

The correlation of the observed effect with theoretical predictions was first attempted on the assumption that the grain density was proportional to the average

rate of energy-loss in silver bromide. However, such an assumption was soon shown to be incorrect, since calculations along these lines predicted a much more marked ratio of plateau to minimum ionization (~1·3) than was actually observed (~1·1). It was then pointed out by Messel and Ritson (1950) that although the energy given to high energy delta-rays is included in the average energy-loss, much of this energy will not be available for blackening the parent silver bromide grain, as the range of such delta rays can easily be greater than the grain size. They therefore proposed that in the energy-loss calculation the maximum transferable energy T (equation (3)) should be replaced by the energy of a delta ray with a range about equal to the grain size; for the emulsions at present in use this is about 2–5 kev. The energy-loss calculated in this way—it is more of a 'probable' than an average energy-loss—shows an increase of 15% between the minimum and the plateau, which is in much better agreement with the observations (figure 13). One ought not, perhaps, to expect too much from such an explanation, since in a standard emulsion nearly 50% of the volume is occupied by silver bromide. Consequently, if a delta ray were to escape from a grain, there is a high probability that it would encounter and expose another, which would then be counted. On the other hand, if only 'blobs' are counted (i.e. no weighting for the number of grains exposed in a delta ray cluster) then the delta ray is virtually ineffective in modifying the grain count once it has left the parent grain—which is effectively what Messel and Ritson's suggestion amounts to.

Probable energy-losses calculated by several authors are compared with the experimental values of grain density in figure 13. As before, the normalization of the various curves is $\mathfrak{J}\to 1$ as $E\to\infty$. It will be seen that while the results are sensitive to the assumptions made regarding the production and re-absorption of Cerenkov radiation, it is quite possible to choose a set of conditions which satisfactorily reproduce the experimental data. However, this apparently good agreement is to some extent fortuitous, since present theories ignore the fluctuations of ionization which, as we shall see in a moment, must play an important part in the mechanism of grain exposure.

A considerable advance towards the proper understanding of the processes involved was made by Brown (1953). He drew attention to the experimental fact that there is a threshold value τ of the energy-release per grain below which the grain is not exposed; what matters therefore is the chance that an energy greater than τ is liberated in a single grain. This chance depends on two factors: first, the path-length of the particle in a grain can vary between wide limits, depending on whether it passes close to the centre, or just grazes the grain; secondly, the energy released in a grain will be subject to the Landau-type fluctuations discussed in §§ 2.3 and 4.1. The probable energy-loss of a minimally ionizing particle in a 0·2 μ silver bromide grain is only about 50 ev. One would therefore expect from West and Rothwell's work with proportional counters (figures 6 and 7) that the fluctuations would be very marked. Moreover, since the value of τ is believed to be around several hundred electron volts (the exact value depending on the type of emulsion and the development) these fluctuations will play a predominant role in the mechanism of grain exposure. Unfortunately, West's observations hardly enable us to draw quantitative conclusions when ξ/I_0Z is as small as about 0·02, which is the case for 0·2μ AgBr grains. Useful experimental evidence could perhaps be obtained by extending his studies to include the case of a counter containing a small pressure of a gas of high atomic number (e.g.

krypton, which has an atomic number intermediate between those of silver and bromide).

It will be obvious from the preceding discussion that attempts to explain the observed effects by calculating most probable energy-losses are—in spite of the often complex nature of the calculations—really no more than first approximations in the treatment of a difficult subject.

5.2. *Crystal Counters and Scintillators*

The post-war development of crystal counters and solid and liquid scintillators placed in the hands of physicists new tools, enabling the energy-loss of particles in condensed media to be studied directly. However, after a few experiments had demonstrated that very little relativistic rise could be observed in silver chloride crystals, and none at all in organic scintillators, interest seems to have waned.

The first of these measurements was made by Whittemore and Street (1949), using a silver chloride crystal. The ionization from mesons of range greater than 112 cm (energy $>1{\cdot}6 \times 10^8$ev), when compared with that from minimally ionizing mesons (3×10^8 ev), showed a definite relativistic increase. The results were quantitatively consistent with the predictions of the Halpern–Hall theory regarding the density-effect correction. Of course, such an 'integral' experiment—in which the high energy group comprises a wide range of meson energies—does not provide a very sensitive method of observing the approach to the Fermi plateau, so that it ought to be worth while repeating this measurement with an improved method of energy determination.

In experiments with scintillators it is not the ionization, but the light output, which is actually measured. It is well known that organic crystals and liquids show a non-linear response for heavily ionizing tracks, but this non-linearity is fortunately negligible for the more lightly ionizing relativistic particles: it was shown by Chou (1952), using pions and protons from the Chicago cyclotron, that most scintillators are nearly linear up to three or four times $I_{\min}$. Mu-mesons passing through anthracene scintillators were found by Bowen and Roser (1951) to show no detectable relativistic increase of ionization above the minimum value. A similar result was obtained by Baskin and Winckler (1953) with a liquid scintillator. However, their measurements had the unsatisfactory feature that the rise in ionization observed at low meson energies was much less rapid than expected, suggesting that saturation might have set in earlier than Chou's results would have indicated.

In the less easily polarizable inorganic scintillators the density effect might be expected to be smaller than in anthracene. This is borne out by a recent result of Bowen (1954): he observed a small relativistic increase of probable ionization, amounting to about 10%, in a sodium iodide scintillator—a value that is in satisfactory agreement with theory.

5.3. *Energy-Loss of Electrons in Thin Foils*

The last group of experiments which will be discussed is that in which the energy-loss of electrons passing through a thin foil is measured with a magnetic analyser. This work is of interest not only as regards the absolute value of the

energy-loss, but also because it affords another test of the theories of energy-loss fluctuation, or 'straggling'. The now classical work of White and Millington (1928), of Madgewick (1928), and others (see Williams 1931 for references) has recently been supplemented by several new investigations. Paul and Reich (1950) measured the energy-loss of 2·8 MeV and 4·7 MeV electrons in foils of beryllium, carbon, iron and lead; Blunck and Leisegang (1950) later compared these measurements with the predictions of their theory of ionization fluctuations. A number of authors (Chen and Warshaw 1951, Jablonski *et al.* 1952, Hildebrand 1953) showed that Landau's theory gave correctly the most probable energy-loss for electrons of energy less than about 2 MeV. Kageyama and Nishimura (1952) also worked with electrons in this energy range, and found that the straggling was greater than Landau's theory would predict. Their work shows good agreement with a study of 624 kev electrons by Kalil and Birkhoff (1953); Blunck and Leisegang's theory satisfactorily explains the results of these two experiments for elements of large atomic number, but predicts widths only half those found experimentally for light elements. A more recent investigation by Hungerford and Birkhoff (1954) shows better agreement for light elements.

Goldwasser, Mills and Hanson (1952) used more energetic electrons (15·7 MeV), and were able to show that the value of the probable energy-loss was in good quantitative agreement with Fermi's value of the density effect correction for beryllium, polythene, aluminium and copper. The discrepancy they reported in the case of gold can be reduced to 6%, according to Warner and Rohrlich (1954), if Fermi's density effect correction is replaced by Sternheimer's more accurate value. An interesting variant of this experiment—giving a direct demonstration of the density effect—has been reported by Goldwasser, Mills and Robillard (1953). Electrons of 15·7 MeV were passed through gaseous absorbers of about one g cm^{-2} thickness, and the probable energy-loss measured. The gases used for this part of the experiment were perfluorocyclobutane and chlorotrifluoroethylene. The experiment was then repeated using the solid (polymerized) forms of these compounds, known respectively as polytetrafluorethylene ('PTFE' or 'Teflon') and 'Kel-F' plastic. In the first case the probable energy-loss was in agreement with Landau's theory, while for the solids it was smaller by an amount equal to the calculated density effect correction.

Acknowledgments

The author wishes to acknowledge with gratitude the help he received during the preparation of this article from discussions with Dr. T. H. R. Skyrme, and from correspondence with a number of authors, particularly Dr. M. M. Shapiro and Dr. P. Budini.

References

Alikhanyan, A., Dadayan, A., *et al.*, 1951, *Doklady Akad. Nauk, S.S.S.R.*, **80**, 37.
Allen, J. R., 1954, *Phys. Rev.*, **93**, 353.
Allison, S. K., and Warshaw, S. D., 1953, *Rev. Mod. Phys.*, **25**, 779.
Bakker, C. J., and Segrè, E., 1951, *Phys. Rev.*, **81**, 489.
Baskin, R., and Winckler, J. R., 1953, *Phys. Rev.*, **92**, 464.
Bassi, P., Bianchi, A. M., and Manduchi, C., 1952, *Nuovo Cim.*, **9**, 861.
Becker, J., Chanson, P., Nageotte, E., Treille, P., Price, B. T., and Rothwell, P., 1952, *Proc. Phys. Soc.* A, **65**, 437.

BEEKMAN, W. J., 1949, *Nature, Lond.*, **163,** 167.
BEHRENS, D. J., 1951, *A.E.R.E. Report T/M* 50 (H.M. Stationery Office, London, Code Number 70-674-0-42).
BETHE, H. A., 1930, *Ann. Phys., Lpz.*, **5,** 325; 1932, *Z. Phys.*, **76,** 293; 1933, *Handb. der Phys.*, **24,** 515.
BETHE, H. A., and ASHKIN, J., 1952, 'Passage of radiations through matter' (Chapter 2 of *Experimental Nuclear Physics*, Vol. 1 (New York: Wiley).)
BETHE, H. A., and HEITLER, W., 1934, *Proc. Roy. Soc.* A, **146,** 83.
BHABHA, H. J., 1937, *Proc. Roy. Soc.* A, **164,** 257.
BLOCH, F., 1933 a, *Ann. Phys., Lpz.*, **16,** 285; 1933 b, *Z. Phys.*, **81,** 363.
BLUNCK, O., and LEISEGANG, S., 1950, *Z. Phys.*, **128,** 500.
BLUNCK, O., and WESTPHAL, K., 1951, *Z. Phys.*, **130,** 641.
BOHR, A., 1948, *Det. Kgl. Dans. Vid. Sels.*, **24,** 19.
BOHR, N., 1915, *Phil. Mag.*, **30,** 581; 1948, *Det. Kgl. Dans. Vid. Sels.*, **18,** 8.
BOWEN, T., 1954, *Nucleonics*, **12,** 48.
BOWEN, T., and ROSER, F. X., 1951 a, *Phys. Rev.*, **82,** 284; 1951 b, *Ibid.*, **83,** 689.
BRADLEY, E. F., 1955, in course of publication.
BROWN, L. M., 1953, *Phys. Rev.*, **90,** 95.
BROWN, W. W., and MCKAY, A. S., 1950, *Phys. Rev.*, **77,** 342.
BROWN, W. W., MCKAY, A. S., and PALMATIER, E. D., 1949, *Phys. Rev.*, **76,** 506.
BUDINI, P., 1953, *Nuovo Cim.*, **10,** 236.
BUDINI, P., and POIANI, G., 1952, *Nuovo Cim.*, **9,** 199.
BUDINI, P., and TAFFARA, L., 1953, *Nuovo Cim.*, **10,** 1489.
CARO, D. E., PARRY, J. K., and RATHGEBER, H. D., 1950, *Nature, Lond.*, **165,** 688; 1951, *Austr. J. Sci. Res.*, **4A,** 16.
CARTER, R. S., and WHITTEMORE, W. L., 1952, *Phys. Rev.*, **87,** 494.
CHANSON, P., TREILLE, P., PRICE, B. T., and WEST, D., 1952, *C. R. Acad. Sci., Paris*, **235,** 1500.
CHEN, J. J. L., and WARSHAW, S. D., 1951, *Phys. Rev.*, **84,** 355.
CHOU, C. N., 1952, *Phys. Rev.*, **87,** 903.
COHN, H. O., ETTER, L. R., and THOMPSON, R. W., 1953, *Phys. Rev.*, **91,** 445 (A).
CORSON, D. R., and BRODE, R. B., 1938, *Phys. Rev.*, **53,** 773.
CORSON, D. R., and KECK, M. R., 1950, *Phys. Rev.*, **79,** 209.
COSYNS, M. G. E., 1936, *Nature, Lond.*, **138,** 284; 1937, *Ibid.*, **139,** 802.
CRANSHAW, T. E., 1951, *Thesis*, University of Cambridge; 1952, *Progress in Nuclear Physics*, **2,** 271 (London: Pergamon Press).
DANFORTH, W. E., and RAMSEY, W. E., 1936, *Phys. Rev.*, **49,** 854.
DANIEL, R. R., DAVIES, J. H., MULVEY, J. H., and PERKINS, D. H., 1952, *Phil. Mag.*, **43,** 753.
DANYSZ, M., LOCK, W. O., and YEKUTIELI, G., 1952, *Nature*, Lond., **169,** 364.
DELLA CORTE, M., RAMAT, M., and RONCHI, L., 1953 a, *Nuovo Cim.*, **10,** 509; 1953 b, *Ibid.*, **10,** 958.
EYEIONS, D. A., OWEN, B. G., PRICE, B. T., and WILSON, J. G., 1955, *Proc. Phys. Soc.* A, **68,** in the press.
FANO, U., 1953, *Phys. Rev.*, **92,** 328.
FERMI, E., 1939, *Phys. Rev.*, **56,** 1242; 1940, *Ibid.*, **57,** 485.
FLEMING, J. R., and LORD, J. J., 1953, *Phys. Rev.*, **92,** 511, 1089.
FOWLER, G. N., and JONES, G. M. D. B., 1953, *Proc. Phys. Soc.* A, **66,** 597.
FRANK, I., and TAMM, I., 1937, *C. R. Acad. Sci., U.S.S.R.*, **14,** 109.
FROST, R. H., and NIELSEN, C. E., 1953, *Phys. Rev.*, **91,** 864.
GARDNER, J. W., 1954, *Phys. Rev.*, **94,** 764.
GHOSH, S. K., JONES, G. M. D. B., and WILSON, J. G., 1952, *Proc. Phys. Soc.* A, **65,** 68; 1954, *Ibid.*, **67,** 331.
GOLDWASSER, E. L., MILLS, F. E., and HANSON, A. O., 1952, *Phys. Rev.*, **88,** 1137.
GOLDWASSER, E. L., MILLS, F. E., and ROBILLARD, T. R., 1953, *Phys. Rev.*, **90,** 378 (A).
HALPERN, O., and HALL, H., 1940, *Phys. Rev.*, **57,** 459; 1948, *Ibid.*, **73,** 477.
HAMMERSLEY, J. M., and MORTON, K. W., 1954, *Biometrika*, **41,** 296.
HAYWARD, E., 1947, *Phys. Rev.*, **72,** 937.

HAZEN, W. E., 1943 a, *Phys. Rev.*, **63**, 107; 1943, *Ibid.*, **64**, 7; 1944, *Ibid.*, **65**, 259; 1945, *Ibid.*, **67**, 269.
HEITLER, W., 1947, *The Quantum Theory of Radiation*, 2nd Edition (Oxford: University Press).
HEREFORD, F. L., 1947, *Phys. Rev.*, **72**, 982; 1948, *Ibid.*, **74**, 574; 1949, *Ibid.*, **75**, 923.
HILDEBRAND, B., 1953, *Phys. Rev.*, **90**, 378 (A).
HINES, K. C., 1955, *Phys. Rev.*, in the press.
HUNGERFORD, E. T., and BIRKHOFF, R. D., 1954, *Phys. Rev.*, **95**, 6.
HÜYBRECHTS, M., and SCHÖNBERG, M., 1952, *Nuovo Cim.*, **9**, 764.
HYAMS, B. D., MYLROI, M. G., OWEN, B. G., and WILSON, J. G., 1950, *Proc. Phys. Soc.* A, **63**, 1053.
INSCH, G. M., 1950, *Phil. Mag.*, **41**, 857.
JABLONSKI, F. E., WALDMAN, B., and MILLER, W. C., 1952, *Phys. Rev.*, **86**, 617 (A).
JAUNEAU, L., and HUG-BOUSSER, F., 1952. *J. Phys. Radium*, **13**, 465.
JELLEY, J. V., 1953, *Progress in Nuclear Physics*, **3**, 84 (London: Pergamon Press).
KAGEYAMA, S., and NISHIMURA, K., 1952, *J. Phys. Soc., Japan*, **7**, 292; 1953, *Ibid.*, **8**, 682.
KALIL, F., and BIRKHOFF, R. D., 1953, *Phys. Rev.*, **91**, 505.
KHARITONOV, V., 1952, *Doklady Akad. Nauk, S.S.S.R.* **86**, 285.
KHARITONOV, V., and BARSKI, A., 1953, *Doklady Akad. Nauk, S.S.S.R.* **93**, 645.
KUPPERIAN, J. E., and PALMATIER, E. D., 1953, *Phys. Rev.*, **91**, 1186.
LANDAU, L., 1944, *J. Phys. U.S.S.R.*, **8**, 201.
LIVINGSTONE, M. S., and BETHE, H. A., 1937, *Rev. Mod. Phys.*, **9**, 245.
LOVATI, A., MURA, A., and SUCCI, C., 1954, *Nuovo Cim.*, **11**, 92.
MADGEWICK, E., 1928, *Proc. Camb. Phil. Soc.*, **120**, 701.
MATHER, R. L., and SEGRÈ, E., 1951, *Phys. Rev.*, **84**, 191.
MCCLURE, G. W., 1953, *Phys. Rev.*, **90**, 796.
MCDIARMID, I. B., 1951, *Phys. Rev.*, **84**, 851.
MESSEL, H., and RITSON, D. M., 1950, *Phil. Mag.*, **41**, 1129.
MICHAELIS, R. P., and VIOLET, C. E., 1953, *Phys. Rev.*, **90**, 723.
MÖLLER, C., 1932, *Ann. Phys., Lpz.*, **14**, 531.
MORRISH, A. H., 1952, *Phil. Mag.*, **43**, 533; 1953, *Phys. Rev.*, **91**, 423.
NIELSEN, C. E., 1941, *Thesis*, University of California; 1942, *Phys. Rev.*, **61**, 202.
NIKITIN, S., 1948, *Zhur. Eksp. teor. Fiz.*, **18**, 577.
OCCHIALINI, G. P. S., 1949, (Como Congress) *Nuovo Cim.*, Supplement, **6**, 377.
PALMATIER, E. D., MEERS, J. T., and ASKEY, C. M., 1954, *Phys. Rev.*, **94**, 766, see also, 1955, *Ibid.*, **97**, 486.
PARRY, J. K., RATHGEBER, H. D., and ROUSE, J. L., 1953, *Proc. Phys. Soc.* A, **66**, 541.
PAUL, W., and REICH, H., 1950, *Z. Phys.*, **127**, 429.
PICKUP, E., and VOYVODIC, 1950, *Phys. Rev.*, **80**, 89.
PRICE, B. T., WEST, D., BECKER, J., CHANSON, P., NAGEOTTE, E., and TREILLE, P., 1953, *Proc. Phys. Soc.* A, **66**, 167.
REYNOLDS, H. K., DUNBAR, D. N. F., WENZEL, W. A., and WHALING, W., 1953, *Phys. Rev.*, **92**, 742.
ROSSI, B., 1952, *High Energy Particles* (New York: Prentice-Hall).
ROSSI, B., and GREISEN, K., 1941, *Rev. Mod. Phys.*, **13**, 240.
ROTHWELL, P., 1951, *Proc. Phys. Soc.* B, **64**, 911.
ROTHWELL, P., and WEST, D., 1950, *Proc. Phys. Soc.* A, **63**, 539.
SCHÖNBERG, M., 1950, *Bull. Cent. Phys. Nucl. Bruxelles* No. 20; 1951, *Nuovo Cim.*, **8**, 159; 1952 a, *Ibid.*, **9**, 210; 1952 b, *Ibid.*, **9**, 372.
SEN GUPTA, R. L., 1940, *Nature, Lond.*, **146**, 65; 1943, *Proc. Nat. Inst. Sci. India* **9**, 295.
SHAMOS, M. H., and HUDES, I., 1951, *Phys. Rev.*, **84**, 1056.
SKRAMSTAD, H. K., and LOUGHRIDGE, D. H., 1936, *Phys. Rev.*, **50**, 677.
SNODGRASS, H. R., 1953, *Phys. Rev.*, **92**, 1059; and private communication.
STERNHEIMER, R. M., 1952, *Phys. Rev.*, **88**, 851; 1953 a, *Ibid.*, **89**, 1148; 1953 b, *Ibid.*, **91**, 256, (*errata*, see *Ibid.*, 1954, **93**, 1434); 1954, *Phys. Rev.*, **93**, 351.
STILLER, B., and SHAPIRO, M. M., 1953, *Phys. Rev.*, **92**, 735.
SWANN, W. F. G., 1938, *J. Franklin Inst.*, **226**, 598.

SYMON, K. R., 1948, *Thesis*, Harvard University (parts of this work are reproduced by Rossi in *High Energy Particles*).
VALENTINE, J. M., and CURRAN, S. C., 1952, *Phil. Mag.*, **43,** 964.
WALKER, W. D., 1953, *Phys. Rev.*, **90,** 234.
WARNER, C., and ROHRLICH, F., 1954, *Phys. Rev.*, **93,** 406.
WEISZ, P. B., and SWANN, W. F. G., 1946, *Phys. Rev.*, **69,** 690.
WEST, D., 1953 a, *Proc. Phys. Soc.* A, **66,** 306; 1953 b, *Progress in Nuclear Physics*, **3,** (London: Pergamon Press).
WHITE, P., and MILLINGTON, G., 1928, *Proc. Roy. Soc.* A, **120,** 701.
WHITTEMORE, W. L., and STREET, J. C., 1949, *Phys. Rev.*, **76,** 1786.
WICK, G. C., 1940, *Ric. Scient.*, **11,** 273; 1941, *Ibid.*, **12,** 858; 1943, *Nuovo Cim.*, **1,** 302.
WILLIAMS, E. J., 1929, *Proc. Roy. Soc.* A, **125,** 420; *Ibid.*, 1931, **130,** 310.
WILLIAMS, E. J., and TERROUX, F. R., 1929, *Proc. Roy. Soc.* A, **126,** 289.
WILSON, J. G., 1946, *Nature, Lond.*, **158,** 414.
WOLFENDALE, A. W., 1953, *Nuovo Cim.*, **10,** 1493; and private communication.
YELISEYEV, G. P., KOSMACHEVSKY, V. K., and LYUBIMOV, V. A., 1953, *Doklady Akad. Nauk S.S.S.R.*, **90,** 995 (English translation: NSF-tr-163, Dept. of Commerce, Washington 25, U.S.A.).

METHODS OF MEASURING STRONG MAGNETIC FIELDS

By J. L. SYMONDS
Department of Physics, The University of Birmingham

CONTENTS

PAGE

§ 1. Introduction 84

§ 2. The phenomenon of magnetic resonance 85
2.1. Introduction 85
2.2. The magnetic moment 86
2.3. The resonance principle 87
2.4. The energy absorption 88
2.5. The line width 88
2.6. The magnetic resonance signal 89
2.7. Proton paramagnetic resonance 91
2.8. Electron paramagnetic resonance 97
2.9. Further applications of the technique 98
2.10. Summary 98

§ 3. The magneto-resistance of bismuth 98
3.1. Introduction 98
3.2. Specimen preparation 99
3.3. Temperature effects 99
3.4. Summary 100

§ 4. The Hall effect 100
4.1. Introduction 100
4.2. Measurement accuracy 101
4.3. Summary 102

§ 5. Peaking strips 102
5.1. Introduction 102
5.2. Measurement accuracy 104
5.3. Other methods of using high permeability material 105
5.4. Summary 105

§ 6. The generator principle 105
6.1. Rotating coil fluxmeters 105
6.2. Vibrating coil fluxmeters 108
6.3. Summary 108

§ 7. Fluxmeter Methods 108
7.1. Mechanical fluxmeters 108
7.2. Fluxmeter coils 110
7.3. Magnetometers 111
7.4. Electronic fluxmeters 112
7.5. Summary 113

§ 8. Forces on current-carrying conductors 114
8.1. Current balances 114
8.2. Particle trajectories: the floating-wire method 116
8.3. Median plane and magnetic centre determination 117
8.4. Summary 117

§ 9. Median plane measurement 117
9.1. Summary 119
§ 10. Miscellaneous methods 119
10.1. The Zeeman effect 119
10.2 Magneto-optical rotation of the plane of polarization 120
10.3. Transmission of light through a colloidal suspension 120
10.4. The magnetron principle 121
10.5. The mercury jet magnetometer 121
10.6. Cathode-ray tube devices 121
10.7. The electrolytic tank 122

§ 11. Conclusion 122

Acknowledgments 122

References 122

Abstract. Descriptions which are briefly theoretical and essentially practical are given of a number of the techniques of magnetic field measurement, used or developed in the period of about twenty years since the last review of the subject. Particular attention is given to the application of the phenomenon of magnetic resonance while other subjects discussed are magneto-resistance, Hall effect, peaking strips, the generator principle and various fluxmeters. The uses which are made of the force exerted on a current-carrying conductor in a magnetic field are given though some of these are not, strictly speaking, field measurements. The measurement of the position of the magnetic median plane of a magnet is discussed. The last paragraphs are devoted to a description of a few methods which do not find very wide application. A list of about 150 references is appended. These are classified in groups under the heading of each particular method.

§ 1. Introduction

It is approximately twenty years since the last reviews were written on the measurement of magnetic fields (McKeehan 1929, Cotton and DuPouy 1932). In this time considerable changes in the techniques available have occurred. The period has been marked by a rapid advance in electron technique, which, to a great extent, has been the result of the wartime development of radar. Furthermore, the construction of the first cyclotron by Lawrence, and all subsequent forms of particle accelerators requiring magnetic guide fields, demanded a precise knowledge of these same guide fields, in magnitude and gradient, to ensure successful acceleration of charged particles to the maximum energy for which the machine was designed. With the advent of the betatron and the synchrotron (both proton and electron), designers were met with the problem of measuring a magnetic field which varied both in space and time. Though some of the older techniques were applicable to these time-varying fields, new ones specific to these fields have now been developed. By far the most significant change in technique has been the discovery of the nuclear and electron paramagnetic resonance. This now provides a high degree of accuracy of measurement which can be achieved in a relatively simple manner and has the further advantage that it gives an answer which is absolute. With the development of these new techniques, there has been a parallel advance in the older styles of measurement, to meet the present need and to make use of improved methods of construction.

The aim of this report is to collect together as much as is practicable of the information which is now available either in the form of published work or internal

laboratory reports. Where possible, this has been sorted into methods which may be treated under separate headings. The literature is mentioned in the text wherever possible but certain papers which are very similar are grouped together in the references. Some of the techniques which are discussed are not in themselves field measurement techniques but are so closely allied to the subject that it was considered advisable to include them. For work carried out earlier than 1933, the reader is advised to consult the work of McKeehan or Cotton and DuPouy, both of which give comprehensive references.

Section 2 is devoted to a fairly detailed account of the nuclear and electron paramagnetic resonance technique with reference to the theoretical aspect of the phenomenon, omitting however any mathematical derivation of formulae which can be readily obtained elsewhere. The emphasis is on the use of the phenomenon as a tool. In §§ 3 and 4, the magneto-resistance and Hall effects are discussed. Comparatively recent studies on these effects in some new materials have shown that they will provide a simple and effective means of making field measurements. Some of the problems in reaching a satisfactory design for an instrument using these effects are mentioned. Highly permeable ferromagnetic materials provide a means of determining when a field goes through zero and this is dealt with in §5, where mention is also made of some essentially low field measuring techniques employing these materials as transformers whose cores undergo varying degrees of saturation. In § 6, the application of the generator principle to field measurement is described and the difficulties found in obtaining accuracies to 0·1% are set out together with methods of achieving this accuracy. This is followed in § 7 by an account of all the methods which integrate the voltage (proportional to dB/dt) obtained from a search coil, giving a value for the difference in field strength in space for static fields, and in time for dynamic fields. Reference is made both to mechanical and to electronic types of fluxmeter. A discussion on the best shape of coil for point measurement is also included. The mechanical force on an electric current has long been a method of achieving high accuracy in field measurement. This is reviewed briefly in § 8, where also it is shown how a stretched current-carrying wire, in a magnetic field, provides a means of demonstrating particle trajectories in that magnetic field. In particle accelerators, the measurement of the position of the magnetic median plane is important and § 9 is given to a study of some methods used in this determination. Section 10 lists six other methods of field measurement which do not find very wide application. In this section, the electrolytic tank technique is briefly mentioned. This has been reviewed recently elsewhere and references are given for the sake of completeness.

§ 2. The Phenomenon of Magnetic Resonance

2.1. *Introduction*

The application of the phenomenon of magnetic resonance to the measurement of magnetic fields probably represents one of the biggest changes in technique which have occurred in this branch of physics in the last decade. The resonance principle was first used by Rabi and his associates (1938, 1939) in the molecular beam method which has developed into a method giving very great precision for the measurement of magnetic moments. A review article by Kellogg and Millman (1946) describes some of the techniques in use in this work.

In 1942, Gorter had made an unsuccessful attempt to detect the effect of

magnetic resonance by the change in the nuclear paramagnetic susceptibility of a substance in a magnetic field. With greatly improved techniques, Purcell, Torrey and Pound (1946) and independently Bloch, Hansen and Packard (1946) introduced the nuclear resonance absorption and the nuclear induction methods respectively. The resonance signals obtained by either of these methods have very narrow line widths and are well adapted to the precise measurement of magnetic fields.

The nuclear resonance techniques were somewhat similar to those used for the investigation of electron paramagnetism by Zavoisky (1945) where it was not necessary to use such sensitive apparatus since the susceptibility due to electron paramagnetism is about a million times larger than the corresponding value for nuclear paramagnetism. This electron resonance technique is suitable for the measurement of fields except that special substances of an organic nature must be used. Previous Reports have dealt in particular with aspects of these techniques (Rollin 1948, Bleaney and Stevens 1953).

In attempting to use these methods for magnetic field measurement, the important feature is that the nuclear resonances occur in the region of radio frequencies for fields from hundreds to thousands of gauss. The electron resonances on the other hand rapidly reach up into the centimetric wavelength region for fields of this order due to the electron magnetic moment being about a thousand times greater than the nuclear moment. For the laboratory where radio-frequency techniques are well known, it is therefore possible to consider electron resonance for measurements up to about 300 gauss overlapping into the region of nuclear resonance for fields up to the order of 20 000 gauss.

2.2. *The Magnetic Moment*

If the magnetic moment vector $\boldsymbol{\mu}$ of a charged particle which possesses a resultant angular momentum is calculated on the basis of the charge being uniformly distributed over a sphere, one finds that

$$\boldsymbol{\mu} = \frac{e}{2Mc}\,\mathbf{p}$$

where $\boldsymbol{\mu}$ is the magnetic moment vector, M is the particle mass, e its charge and $\mathbf{p}$ its angular momentum vector with c the velocity of light. In general, this picture is too simple and it is necessary to multiply by a dimensionless constant g to reach agreement with experiment, so that

$$\boldsymbol{\mu} = g\,(e/2Mc)\,\mathbf{p} = \gamma\mathbf{p}.$$

For the case of a free electron, this constant g is very close to 2 and is often known as the Landé g-factor or as the spectroscopic splitting factor. It should not be confused with the gyromagnetic ratio γ which is the ratio of the magnetic moment to the angular momentum.

If this charged particle is placed in a constant external magnetic field $\mathbf{B}_0$, the angular momentum vector $\mathbf{p}$ will precess around the direction of $\mathbf{B}_0$, with an orientation such that its component along the field assumes values equal to $mh/2\pi$. In this, m can only take values equal to or between $\pm I$, where I is the spin of the particle. Thus there are $2I \pm 1$ different energy levels which correspond to the

Zeeman splitting of levels in a magnetic field. The difference between the two adjacent levels is

$$\Delta E = h\nu_0 = g(eh/4\pi Mc)B_0$$

or

$$\omega_0 = 2\pi\nu_0 = g(e/2Mc)\, B_0 = \gamma B_0$$

where ν_0 = the Larmor frequency of the precession and γ = the gyromagnetic ratio. In the cases which need consideration here $I=1/2$, i.e. for protons and electrons. As a result, there are only two possible levels, one with the component of the magnetic moment vector along the field, μ (commonly known as the magnetic moment), parallel to the field $\mathbf{B}_0$, and the other with it antiparallel. It should be noted that the Bohr magneton for electrons (namely $eh/4\pi mc$) is 1836 times greater than the corresponding unit for protons, the nuclear magneton.

2.3. *The Resonance Principle*

With the particles in a constant magnetic field B_0, let us suppose that we supply a small alternating field perpendicular to the direction of B_0. In the classical picture, this can be resolved into two circularly polarized components, one B_1 rotating in the same sense as the precessing system and another in the opposite sense. The former will exert a torque

$$\mathbf{T} = [\boldsymbol{\mu} \times \mathbf{B}_1]$$

and if the rotating field frequency ω is identical with the Larmor frequency ω_0, the angle θ between $\mathbf{B}_0$ and $\boldsymbol{\mu}$ will change and a reorientation will occur (figure 1).

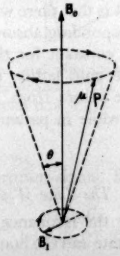


Figure 1. A vector representation of the precession of the angular momentum vector $\mathbf{p}$, and hence the magnetic moment vector μ, about the direction of the magnetic field $\mathbf{B}_0$. The small rotating field $\mathbf{B}_1$ is shown rotating in the same sense as the precessing vector and this will tend to increase the angle θ.

However, if $\omega \neq \omega_0$, the average effect will be zero. The counter-rotating component exerts no average effect on $\boldsymbol{\mu}$, either with $\omega = \omega_0$ or with $\omega \neq \omega_0$.

In the quantum picture of the same process, a quantum $h\nu$ will change the orientation only in the integral multiple of the correct value ΔE for the difference between the energy levels. The probability of a parallel spin absorbing energy is enormously increased when $\omega = \omega_0$ and we can say that a resonance has occurred.

2.4. *The Energy Absorption*

Consider a sample containing the required particles with spin 1/2 at about room temperature. In the absence of magnetic field, the number of particles N_a with spin $+1/2$ will equal the number N_b with spin $-1/2$ with no difference in the energy levels. In the presence of a magnetic field B_0, the levels are split and the energy level difference is $\Delta E = \gamma h B_0/2\pi kT$. This difference is much less than the thermal energy kT so that the distribution of particles between the two spin states $+1/2$ in the direction of, and $-1/2$ opposing, the field will be given by the Boltzmann law

$$\frac{N_b}{N_a} = \exp(-\Delta E/kT) = \exp(-\gamma h B_0/2\pi kT)$$
$$\simeq 1 - \gamma h B_0/2\pi kT.$$

For protons in a magnetic field of 10^4 gauss at room temperature, this ratio is different from unity by about 7×10^{-6}. It is this small excess of particles with spins in the direction of the field which makes it possible to obtain an absorption of power.

This consideration alone is not sufficient to explain the absorption which is normally observed. It is possible to imagine a condition where N_b is so much increased by the action of the external alternating field B_1 that N_a tends to the value N_b. When this condition is finally reached, almost as much energy is handed back to the field as is taken from it and the resultant energy absorption would fall below the limit of detection. This phenomenon is known as saturation.

The lattice in which the affected particles reside must undergo thermal motions, as for example the Brownian movement in liquids, and an exchange of energy can occur between the magnetic energy of the particles and the kinetic energy of thermal motion. The mean result is that there will be a release of magnetic energy to the thermal motion and a corresponding absorption of energy from the applied radio-frequency field. The time constant for the system to adjust itself to the new conditions is known as the spin–lattice relaxation time T_1. This time constant T_1 can have very diverse values from 10^{-4} second to several hours in nuclear resonance (Purcell 1951) while in paramagnetic resonance it is generally shorter than 10^{-4} second.

2.5. *The Line Width*

The line width associated with the resonance phenomenon is connected with the lifetime of the spins in a given state and the homogeneity of the applied magnetic field. The time which the spin can spend in a given state can be changed in several ways, principal among which are: (i) the actual splitting of energy levels as a result of the perturbation introduced into the system, (ii) the interaction of fields local to the particular particles causing each to be in a different field and thus producing a spread in a given energy level. This process, in fact, is associated with the finite lifetime of a given state arising from the relaxation mechanism. These interactions cause the particles to lose their memory for the particular phase of precession. The time taken for the particles to lose coherence in phase is called the spin–spin relaxation time T_2.

In a perfectly homogeneous magnet, Bloch has shown that the half-width of the resonance is then $1/T_2$ but this cannot be achieved in general as inhomogeneities

in the magnet can cause widths to exceed this value by as much as several orders. The line width $\Delta\omega$ in a magnet which has inhomogeneities over the sample, ΔB, is given by

$$\Delta\omega = 2/T_2 + \gamma\Delta B.$$

2.6. *The Magnetic Resonance Signal*

Bloch (1946) has shown in his phenomenological theory that the susceptibility $\chi=\chi'-i\chi''$ of the sample undergoing resonance can be written in terms of the volume susceptibility χ_0, γ, T_1, T_2, B_0, B_1, and the difference of the applied frequency ω from the true resonant value ω_0. The use of complex notation in the analysis of a dispersive system is common and analogy should perhaps be drawn with the case of anomalous dispersion where the index of refraction and the absorption coefficient are treated as the real and imaginary part of a complex refractive index. The out-of-phase component χ'', in the complex susceptibility, gives the absorption signal while the in-phase component χ' gives the dispersion signal, figure 2 showing

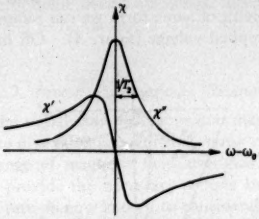


Figure 2. The variation of the real and the imaginary components χ' and χ'' of the magnetic susceptibility near resonance. The half-width of the resonance is shown as the reciprocal of the relaxation time T_2.

how χ' and χ'' will vary with $\omega-\omega_0$.

If we consider only the absorption signal, Bloch has shown that at resonance

$$\chi''=\frac{\gamma T_2\chi_0 B_0}{1+(\gamma B_1)^2T_1T_2}.$$

The contribution of χ'' to the absorbed power is, on average,

$$P=\omega B_1^2\chi''.$$

Now if $(\gamma B_1)^2T_1T_2\ll 1$, the absorbed power will increase with B_1^2 but for $(\gamma B_1)^2T_1T_2>1$, there will be a condition where the absorbed power is constant, i.e. saturation occurs.

A sample of the requisite material is placed in a coil of inductance L through which a radio-frequency current i passes, generating the perturbing field B_1. Provided the coupling to the generator is weak, we can consider the sample circuit as supplied by a constant current generator (figure 3). The inductance L can be

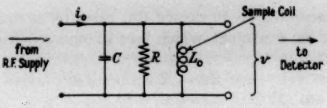


Figure 3. The equivalent of the magnetic resonance unit which is loosely coupled to the radio-frequency supply.

written as $L=L_0(1+4\pi\chi q)$ where q is the filling factor for the sample in the coil.

At resonance, $\omega C = 1/\omega L$ and the admittance becomes

$$Y=\frac{1}{R}+i\omega C\left[1-\frac{1}{\omega^2 L_0 C(1+4\pi\chi q)}\right]$$

or, since $4\pi\chi q \ll 1$

$$Y\simeq \frac{1}{R}+\frac{4\pi i\chi q}{\omega L_0}.$$

Thus the real part of the admittance is a function of the absorption χ'' and the imaginary part a function of the dispersion χ'. The voltage across the sample coil is $v=v_0(1-4\pi i\chi qQ)$ where $v_0=i_0R$ and $Q=R/\omega L_0$.

If we use a bridge circuit of some form we can remove most of that part of v which comes from the applied voltage (figure 4). Off magnetic resonance, there

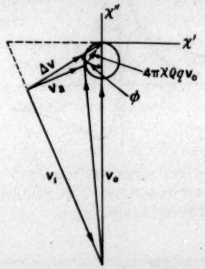


Figure 4. Vector diagram of the two opposing voltages, v_0 and v_1, from a bridge circuit, showing the voltage resulting from the magnetic resonance also: $4\pi\chi Qqv_0$.

will be a relatively small signal Δv received by the amplifier, whereas on resonance, it will receive $v_a=\Delta v-4\pi i\chi qQv_0$. The value of the second term is small compared with the first so that the signal voltage v_r seen on detection is the real part of v_a, with ϕ, the difference in phase between v_0 and Δv:

$$v_r\simeq 4\pi Qq|v_0|[-\chi''\cos\phi+\chi'\sin\phi].$$

This means that the signal is a pure absorption signal only if $\phi=0$ or π, while it is a pure dispersion signal if the bridge is balanced in amplitude ($\phi=\pm\pi/2$).

For protons at room temperature, the nuclear susceptibility χ_0 per unit volume is given by the Langevin formula

$$\chi_0=\frac{N\mu_p{}^2}{kT}=3{\cdot}4\times10^{-10}\text{ erg gauss}^{-2}$$

where $\mu_p = 1{\cdot}41 \times 10^{-23}$ erg gauss^{-1} and N=number of protons per cm^3 in the sample. Let us take the magnetic field inhomogeneities as the limiting factor in the line width and have $T_2 \simeq 10^{-4}$ sec, equivalent to about 1 gauss over the sample of 1 cm^3. In the region where $(\gamma B_1)^2 T_1 T_2 \ll 1$, taking $|v_0|$ as one volt, q as 0·9 and Q as 100, the signal would be about one millivolt for $B_0 \simeq 1000$ gauss.

To obtain optimum signal for given conditions, the value of T_1 and that part of T_2 due to T_1 can be changed by the admixture of a salt producing paramagnetic ions in the water sample used for giving proton resonance. The interaction of the nuclear moment with its surroundings is increased and as a result the relaxation time T_1 is decreased and the line width increased somewhat. Figures for given admixtures can be found in the work of Dickinson (1951), Bloembergen (1948) and Gabillard (1952). They show that T_1 is inversely proportional to the concentration of ferric ions, a relaxation time of $6{\cdot}5 \times 10^{-4}$ second being reached for 10^{20} ferric ions per cm^3. The advantage in adjusting T_1 lies in the ability to fulfil the condition $(\gamma B_1)^2 T_1 T_2 \ll 1$, in given circumstances, enabling a higher B_1 to be used before saturation occurs.

Bloembergen, Purcell and Pound (1948) give an expression for the maximum signal-to-noise ratio which can be achieved under a given set of conditions. This always assumes that microphonic noise and outside interference can be eliminated and that the magnetic field to be measured is not too inhomogeneous as is often the case.

2.7. *Proton Paramagnetic Resonance*

The magnitude of the signal obtained by proton magnetic resonance has been shown to be sufficient to give adequate signal-to-noise ratios with straightforward circuits over a large range of magnetic field strengths. The use of protons in water or mineral oil to provide the necessary sample has been adopted generally since the gyromagnetic ratio is now known to considerable accuracy:

$$\gamma_H = (2{\cdot}67523 \pm 0{\cdot}00006) \times 10^4 \text{sec}^{-1}\text{gauss}^{-1}.$$

Due to the long relaxation time of protons in water ($T \simeq 2{\cdot}3$ sec) it is usual to add some paramagnetic ions as discussed in § 2.6, unless great precision is required. For normal measurements the change in the field at the nucleus can be ignored as it represents a correction of about 1 part in 10^5. Similar corrections apply to the measurements made with mineral oils.

(*a*) *Modulation.*

For the observation of the complete absorption or dispersion signal, it is necessary to sweep through the resonance. This can be done either by modulation of the magnetic field over a range larger than the line width or by modulation of the frequency. The literature seems to indicate that the former method is more usual and less troublesome. In the use of frequency modulation there appears to be some difficulty in assessing the true frequency of the resonance. The frequency modulation method has been used by Briggs (1950) in apparatus for the measurement and control of the magnetic field of the Illinois cyclotron.

(*b*) *Signal detection.*

There are now in existence a number of experimental arrangements which can be used to detect the resonance. Fundamentally, they all rely on placing the

sample in a coil which forms part of a tuned circuit whose impedance suffers a change when the resonance occurs. Provided the generator is sufficiently stable in frequency and amplitude, a resonance signal can be detected on the coil voltage if the coil is supplied with constant current by way of a large resistance from the signal generator (Rollin 1948) or if loose coupling to the sample circuit is made by a loop about the coil (Richter, Humphrey and Yost 1954; see also figure 5

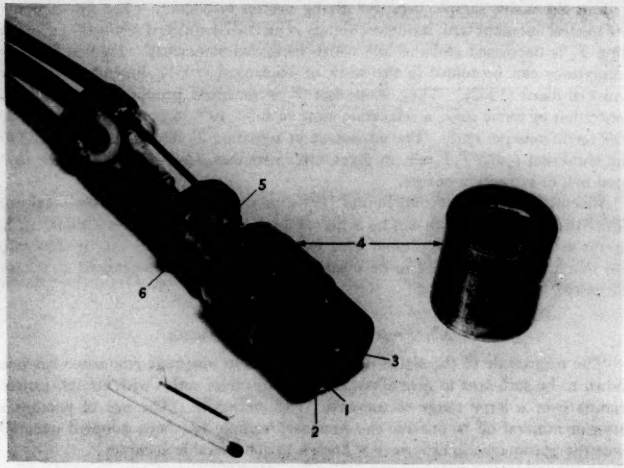


Figure 5. The nuclear resonance head used for measurements on the Birmingham cyclotron magnet. A spare sample coil, wound on a glass tube containing the necessary liquid, is shown alongside a match for comparison. The modulation coils, 2, are on either side of the sample coil, 3, and the coupling loop, 1. The germanium diode detector, 5, and the tuning condenser, 6, are behind the coil, all inside the main housing, 4. (*By courtesy of H. Shaylour.*)

of Birmingham equipment). These have the advantage of simplicity, one control being needed to set up the circuit.

(*c*) *Balance circuits.*

Thomas and Huntoon (1949) have developed a circuit which they have designated an amplitude bridge (figure 6). As the name indicates, this arrangement will detect only amplitude and not phase changes in the radio-frequency voltage on the coil. In consequence, only the absorption signal is registered. To reduce the amplitude noise from the generator to a minimum, the rectified output from the sample circuit and that from a dummy tuned circuit are balanced against each other. The resultant output will consist of the required signal when the resonance occurs plus noise generated in the circuit and the rectifiers. The balance to reduce amplitude noise does not necessarily apply for any frequency modulation noise at the generator output. Although crystal diodes generate

more noise than pure tungsten filament diodes, the former were found to work quite satisfactorily at fields above a few thousand gauss.

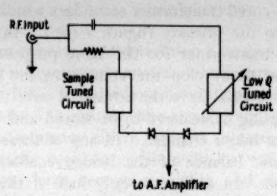


Figure 6. A schematic diagram of the amplitude bridge for nuclear magnetic resonance detection: *after* Thomas and Huntoon.

By far the most effective technique for achieving a balance involves the use of some form of radio-frequency bridge. The reduced off-resonance voltage applied to the amplifier and detector system enables a lower noise level to be maintained. Purcell *et al.* (1946) proposed such an arrangement in their original investigations and, in general, the high-frequency bridges used are adaptations of this method. The two arms of the bridge are identical in construction and are loosely coupled together by some system for changing the phase of the voltage

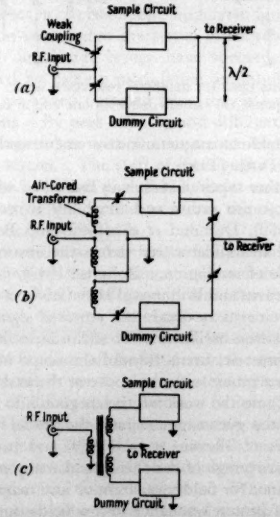


Figure 7. Various radio-frequency bridges used in detecting nuclear magnetic resonance: (*a*) *after* Purcell, (*b*) *after* Lindström, (*c*) used in Birmingham.

from one arm by 180°. Purcell used a difference in cable length of $\lambda/2$ between the connections from the two arms of the bridge (figure 7 (*a*)). Lindström (1951) uses a centre-tapped air-cored transformer secondary winding, the radio-frequency voltage being applied to the primary (figure 7 (*b*)). In Birmingham, we have used a Mumetal-cored transformer for the same purpose (figure 7 (*c*)). This does not require tuning to develop the voltage on the secondary. The latter two methods for phase reversal have the advantage over the original method that no long lengths of coupling cable have to be stored and they can cover a wider frequency range without major changes. In any of these systems adjustment of the phase and amplitude balance of the bridge enables observation of either absorption or dispersion. In each case only half of the true resonance signal appears at the output.

(*d*) *Other circuits.*

A bridge-T rejection circuit has been used with success in place of the above double arm type (Waring *et al.* 1952). Two forms of the bridge-T network can be used (figure 8), the circuit with one end of the coil at earth potential appearing

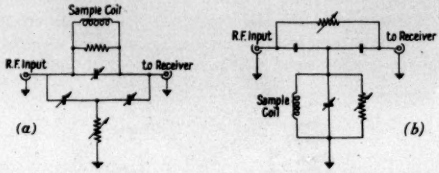


Figure 8. Bridge-T circuits used for magnetic resonance work: (*a*) sample coil in bridge with no earth point, (*b*) sample coil with one end at earth potential.

to be more readily adaptable to magnetic measurement work. Bridge-T networks are treated in detail by Tuttle (1940).

There are various other types of circuit in use all of which appear to provide very adequate signal-to-noise ratios and flexibility in control. Pound (1947), Pound and Knight (1950), Dumond *et al.* (1952) and Beljers (1953) place the specimen in the coil of an oscillator and detect the absorption of power by the change in the amplitude of oscillation. Roberts (1947), Gutowsky *et al.* (1953), Hopkins (1949), Zimmerman and Williams (1948), Grivet *et al.* (1953) use autodyne and super-regenerative circuits to obtain the required signals while Knoebel and Hahn (1951) use a transitron oscillator. To mention in detail all the advantages and disadvantages of these circuits is beyond the scope of this report and a list of references is given in a group with the above at the end of the paper.

To obtain an insight into the work which has gone into achieving the accuracy in the measurement of the gyromagnetic ratio the reader would be well advised to investigate the papers of Thomas *et al.* (1950) and Lindström (1951). The latter gives an excellent coverage of the theory and many practical instructions in the use of proton resonance for field measurement and magnetic field control.

Stretching the use of proton resonance to low fields can prove exasperating in the extreme and its use in fields as low as 35 gauss for stabilizing a β-ray spectrograph magnet is analysed in detail by DuMond *et al.* (1952). Other circuits

for field stabilization by proton magnetic resonance are given by Thomas (1952) and Packard (1948).

(*e*) *The ultimate accuracy.*

The one measurement necessary to determine a magnetic field by this method is that of the frequency at magnetic resonance and the accuracy in this determination can easily reach 1 part in 10^5. Standard oscillators, using quartz crystal oscillators for comparison, may be used to estimate the signal frequency. The two outputs are mixed and the beat frequency either reduced to zero, or estimated by comparison with a standard audio-oscillator. Since the last war, considerable numbers of heterodyne wavemeters have become available, and enable a direct measurement to be made with the one unit to an accuracy of the order of 1 part in 10^4. These units have an oscillator with continuously variable frequency which may be checked at the harmonics of a 100 kc/s quartz crystal oscillator. The frequency of the signal to be measured is heterodyned with a harmonic of the continuously variable oscillator, the dial reading giving the true frequency from an interpolation chart. All this assumes, however, that the frequency of the oscillator has been set exactly for magnetic resonance. A fundamental part of the measurement is therefore finding a satisfactory method for the estimation of the true resonance point.

If the modulating field is small compared with the line width it is possible to use a meter to show the signal magnitude at any frequency, or the output can be displayed on a chart recorder as the frequency is varied. This method unfortunately presupposes a considerable understanding of the signal shape unless steps are taken to ensure that the signal is entirely absorption. Thomas *et al.* (1950) with their amplitude bridge have succeeded in doing this. The entire signal is displayed by using a modulation amplitude greater than the line width and an oscilloscope sweep obtained from the sinusoidal voltage applied to the field modulation coils. By peak clipping and differentiating this same voltage and mixing the resulting pulses with the resonance signal, vertical lines were made to appear on the screen. The shift of the signal could be readily observed relative to these markers. A small pair of Helmholtz coils, calibrated by shifting the frequency and adjusting the current to return the signal to its first position, could be used to give interpolation fields.

Lindström (1951) describes another method which involves the photography of oscillograph traces showing the resonance signal several times per trace, under conditions where the modulation is greater than the line width. A series of photographs is taken with different frequency settings and the linear differences between the signals measured. The differences between these signals, $a_{23}-a_{21}$ in figure 9, are plotted as a function of the frequency, the intercept on the frequency axis giving the true resonance point. He uses this technique in the comparison of magnetic moments of different substances. Beljers (1953) uses a somewhat different technique. The modulation is again much greater than the line width and several signals are made to appear on the oscilloscope trace (figure 9). If the frequency is too low and the current in the modulating coil is reduced, the signals 1 and 2 will move together and finally disappear. When the frequency has been accurately adjusted to the value which coincides with the Larmor precession frequency for the basic magnetic field B_0, the signals will remain stationary, though the signal width will appear to increase.

(*f*) *Use in inhomogeneous fields.*

With the need for precision measurements in particle accelerators, the question of inhomogeneous fields arises. Here the accuracy will be limited by the enforced

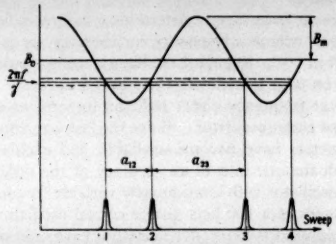


Figure 9. For the production of magnetic resonance signals, the modulating field B_m brings the main field B_0 to the resonance condition $2\pi f/\gamma$ at points 1, 2, 3 and 4. When these points occur at equal intervals (i.e. as B_m goes through zero), the frequency is set correctly for resonance at the field B_0.

use of paramagnetic solutions with broad resonance lines. It seems to be generally accepted that a line width of approximately one gauss will provide a sufficiently accurate answer in these applications. This means that the more inhomogeneous the field to be measured, the smaller must the sample be. There is a compensating factor in the magnitude of the signal since here the more inhomogeneous fields are usually greater in absolute magnitude and the signal is proportional approximately to the three-halves power of the field. Extermann *et al.* (1953) consider that the resonance method is still convenient for measurements in field gradients of 50 gauss per mm.

(*g*) *Use in rising fields.*

In synchrotrons, it would seem reasonable to use the proton resonance as a means of measurement since the sweeping field is automatically present. However, the rate of sweep is usually high, of the order of 15 kilogauss per second, and the signal is complicated by transient effects. These are the result of the frequency of nuclear precession changing with the field and the radio-frequency signal, resulting from the changing precession, beats with the applied radio-frequency signal, causing maxima and minima. The amplitude of these maxima and minima decays as the nuclei get out of phase with one another, and this decay is therefore related to T_2, or the combination of T_2 and the effect of an inhomogeneous field $1/\gamma\Delta B$.

Unless the shape of this signal is fully understood, error can arise from the choice of the point on the signal at which the true resonance is assumed to occur. Useful information on this topic is contained in the theoretical paper of Jacobsohn and Wangsness (1948) which makes it possible to calculate some figure for the delay between the true resonance point and some fixed point on the signal, say the first maximum. Gabillard (1952) gives useful experimental data on these complicated signals.

2.8. *Electron Paramagnetic Resonance*

The principle is identical with that for nuclei but it is applied in this case to organic substances which contain ' free radicals ' with unpaired electron spins. The observation of the paramagnetic resonance in such compounds was reported by Kosirev and Salekov (1947) and Holden, Kittel, Merritt and Yager (1949, 1950). The particular unpaired electrons in these compounds behave as though they are almost free and the g-value is very close to the free spin value (Holden *et al.* 1951, Townes and Turkevitch 1950). Since the gyromagnetic ratio for the electron is some 660 times greater than that for the proton, very much larger signals will result. So far, the best material for measurements on magnetic fields is diphenyl picryl hydrazyl, since in its polycrystalline form it provides the sharpest line yet known, with a half-width of 1·4 gauss and a g-value of $2{\cdot}0038 \pm 0{\cdot}0002$. Using the figures of Taub and Kusch (1949) and Thomas *et al.* (1950) this gives $f = (2{\cdot}804 \pm 0{\cdot}001)B$ Mc/s with B in gauss. The relaxation time is very short in this substance and it is possible to achieve either absorption or dispersion signals of simple form even with rates of rise of field in the region of 30 kg sec^{-1}.

(*a*) *Signal detection.*

The resonant frequency is much higher for a given B than in the equivalent proton resonance and as a result the same kind of apparatus is only practicable to about 300 gauss. Easily recognized signals may be obtained below 10 gauss. Generally, the sample needed is less than 1 cm^3 of the polycrystalline solid.

In Birmingham, two units working at 550 and 1000 Mc/s have been constructed to provide monitor signals for the magnetic field of the proton synchrotron in the region of the injection magnetic field. These rely on the rising field of the synchrotron for the necessary modulation. An accuracy of about 1 part in 500 can be achieved in this magnetic field range. Another unit working at about 40 gauss (110 Mc/s) using a 50 c/s modulation field is used for checking the remanent field between magnet cycles every 10 seconds. It has been possible to stretch the technique of using the bridge-T network, as shown in figure 8 (*b*), to this frequency and the output is directly detected by a germanium crystal diode. Signals of a millivolt are readily obtained with fractions of a cubic centimetre of the organic compound.

For higher fields, waveguide and cavity techniques become more applicable (Roberts and Jones 1951). In the measurement of other g-values close to 2, the organic substance quoted is often used as a standard of comparison. This is treated in a previous Report by Bleaney and Stevens (1953) where further references on the use of paramagnetic resonance at higher fields may be obtained.

(*b*) *Preparation of a sample.*

Consideration of the difficulty experienced in obtaining the necessary material prompts the author to add a chemical recipe for the sake of progress in physics. 1:1-diphenyl-2-picryl hydrazyl (Goldschmidt 1920, 1922, 1924) was prepared in the following manner: 1:1-diphenyl-2-picrylhydrazine was formed (94% yield) by the condensation of *asym*-diphenylhydrazine (which is commercially available) and picryl chloride. Oxidation of this with lead peroxide in chloroform gave the hydrazyl (74% yield). The compound forms violet plate crystals from chloroform–ether. The author is indebted to the Chemistry Department, University of Birmingham, for their help.

2.9. *Further Applications of the Technique*

Emphasis has been placed in the preceding discussion upon the use of resonance techniques as standards of comparison employing one unit. Both the proton and the electron resonance methods lend themselves to the measurement of point-to-point variations in a field. By the simple expedient of supplying two resonance heads from the one oscillator, modulating the field with the same alternating current passing through similar coils and displaying the signals on a double beam oscilloscope, a very easy and accurate determination of field variations in space can be made. One head is left at some fixed reference point in the field while the other is moved over the region to be investigated. The problem of accurate control of the field is avoided in this way provided the overall field fluctuations do not move either signal out of the range of the modulating field.

In the case of rising fields, the procedure is to display the signals on a linear timebase with known sweep rate. Previous or simultaneous measurement of dB/dt at the resonance instant enables B to be estimated from the time interval between the signals.

2.10. *Summary*

The magnetic resonance technique provides a very useful and precise means of field measurement. Its use has become very widespread and there are now available some commercial units using nuclear resonance which are designed for a wide range of frequencies and fields (1–100 Mc/s or 200–10 000 gauss).

As a means of determining the mean value of an extensive magnetic field, the method has its limitations. It is, perforce, a point measurement and, unless the field is very homogeneous, it is not very practicable or reasonable to expect a great deal from a few measurements at spot points, but this is a difficulty which arises in all point measurement techniques.

For the accurate control and stabilization of magnetic fields it has proved to be a reliable standard arrangement suitable for general use. It has the enormous advantage of providing its own standard for assessing the field. In the past, current measurement has been used but this need not bear a unique relationship to the field strength.

§ 3. The Magneto-Resistance of Bismuth

3.1. *Introduction*

Although it is possible to measure a change of electrical resistance in many materials when they are subjected to a magnetic field, the comparatively large change in bismuth makes it possible to contemplate using it for magnetic field measurement. The 'bismuth spiral' as a measuring tool was one of the earliest and best known applications of the magneto-resistance of bismuth but it was discarded for the most part because the resistance depended very greatly on the contact made between the bismuth and the external leads. Furthermore the unreliable and poor cohesion of the specimen made the ' spiral ' a rather fragile structure (Kapitza 1928). Another great disadvantage was the marked change induced by mechanical stress in the specimens (Allen 1932). In fact, perusal of the literature shows considerable information on the attempted use of this phenomenon. Some of these references are listed.

3.2. *Specimen Preparation*

The preparation of the specimen was therefore the major initial problem in any attempt to use the phenomenon and, in the last few years, some attention has been devoted to making wires of bismuth in a sufficiently ductile form to permit a small coil to be wound. Conn and Donovan (1951) describe a fairly simple method for making these bismuth wires in which they draw down thin soda glass fibres previously filled with molten bismuth. The final diameter of the fibre is about $\frac{1}{2}$ mm. Filaments prepared in this way have the glass subsequently removed with hydrofluoric acid, leaving a very soft, thin wire. Resistances of between ten and a hundred ohms per cm are realized. Although the bismuth is under considerable compression in the glass due to the expansion of bismuth on cooling, this is relieved on etching and the anisotropy of resistance with orientation in the field is largely removed in the process of winding which probably causes some modification of the crystalline form. Coils with a resistance of about 200 ohms have been wound on formers of $\frac{1}{2}$ mm diameter with 3 to 4 mm winding length. In the final specimens, Conn and Donovan observed anisotropies which, in general, were about 3% of the change of resistance at any given field. Unetched specimens would show considerably greater anisotropy (Donovan and Conn 1949).

3.3. *Temperature Effects*

Another rather unfortunate property of bismuth is the large resistance–temperature effect which it shows. Reference to tables (see *International Critical Tables*, Vol. **6**, 422, Table 16) demonstrates the obvious advantage of working at low temperatures, the fractional resistance change being about twenty times greater at liquid nitrogen temperature than at room temperature. Where it is important to measure fields below 1000 gauss by this method, liquid nitrogen temperatures are almost essential since, at room temperatures, there is only an effect of about 1% in going from zero to 10^3 gauss.

Keller (1953) has used freon 12 (about -28°c) as refrigerant on non-inductive coils formed of ductile bismuth wire (extruded to 0·005 in. diameter) and with a total zero-field resistance of about 200 ohms. For fields above 5000 gauss, he finds that the resistance is a linear function of the applied field to within 0·1%. The relationship in this region is determined experimentally at two points by comparison with a nuclear resonance detector. For fields below 5000 gauss the exact shape of the resistance–field curve is needed, but if it is determined for one probe over this range of field then all others may be calibrated by a similarity transformation.

Both Conn and Donovan, and Keller agree that these coils show no erratic variations in resistance. Keller states that, over an eight hour period, it was possible to repeat absolute measurements to better than 1 part in 1000. This considerable improvement in stability seems to be due partly to the improvement in the structural form of the bismuth and also to the better electrical connection. 'Wood's' metal seems to provide an adequate connection though it must be skilfully applied to avoid growth in the grain size in the bismuth wire due to overheating. Keller also mentions the use of printed-circuit silver paint but this apparently shows an ageing effect at low temperatures. The method has been applied to both a.c. and d.c. measurements, though for a.c. fields there is evidence of a lag in the resistance of the bismuth (König 1908, Heurlinger 1916, Smith 1939) when the field is applied.

3.4. *Summary*

The measurement of resistance is both simple and rapid so that the specimens of bismuth as prepared seem to provide a very useful method of at least comparative field measurements as long as the temperature variation is known or the temperature is held constant. Experience in one instance has also shown that precise measurements can be made using bismuth coils at low temperatures. The problem of specimen preparation appears to have been solved though the technique demands some skill in producing specimens of known resistance per unit length and reliable polycrystallinity.

§ 4. The Hall Effect

4.1. *Introduction*

In 1879, Hall found that when a strip of metal which was carrying current was placed in a magnetic field, a potential difference appeared across the strip. The effect is usually measured by passing a current along a thin strip of material placed in a magnetic field which is normal to the surface of the strip. The resultant voltage produced across the strip between X and Y (figure 10) is given by

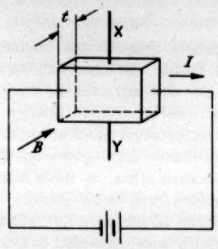


Figure 10. Diagram showing the essential features of the production of the Hall voltage across a specimen.

$$V_h = R_H IB/t \text{ volts} = \frac{R_H}{\rho}\frac{VBl_H}{l_I}$$

where R_H = the Hall coefficient in v cm a^{-1} gauss^{-1}, I = current in the specimen in amps, B = magnetic field in gauss, t = strip thickness in cm, V = voltage applied to the specimen, l_I = specimen length along current, l_H = specimen width in Hall voltage direction.

The appearance of this voltage across the strip is the outward sign of the force exerted on the current carriers by the magnetic field. The tendency to force them sideways in the specimen results in the formation of a space charge along the sides of the specimen and a consequent transverse electric field is set up which is just sufficient to balance the effect of the magnetic field.

The current carriers may, in some cases, be 'holes' (in p-type conductors) or they may be electrons (n-type conductors). In hole conductors, the Hall voltage will be negative, i.e. in figure 10 Y would be positive with respect to X, while in electron conductors this voltage is reversed.

To achieve conditions where the Hall voltage is sufficiently large to be used for the measurement of magnetic fields without undue complication, the ratio of the Hall coefficient to the resistivity must be as large as possible since, for a given sample size, magnetic field and driving voltage, V_H = constant (R_H/ρ) where ρ = specific resistivity. However, for ease of application, the Hall coefficient itself should be large otherwise prohibitive main currents would have to be used. The table gives some of the figures for various substances at room temperature. Bublitz (1938) has used bismuth but this material suffers from lack of cohesion, making it difficult to obtain stable conditions. The favourable ratio of germanium coupled with the surprisingly large Hall coefficient makes it particularly suitable for magnetic field measurements. There is the added advantage that germanium is a refractory

Material	R_H (v cm A^{-1}gauss^{-1})	ρ (ohm cm)	R_H/ρ
Germanium (n-type)	$3{\cdot}6 \times 10^{-4}$	10	$3{\cdot}6 \times 10^{-5}$
Bismuth	1×10^{-8}	$1{\cdot}1 \times 10^{-4}$	9×10^{-5}
Copper	$5{\cdot}2 \times 10^{-13}$	$1{\cdot}7 \times 10^{-6}$	3×10^{-7}
Indium antimonide	$2{\cdot}4 \times 10^{-5}$	3×10^{-2}	8×10^{-4}

material (melting point 958°c) from which convenient plates may be cut with a diamond saw. These may be ground subsequently to the required size. Pearson (1948) indicates that samples as thin as 0·004 in. have been prepared but the usual thickness used by him was about 0·04 in., the plate being 0·5 in. × 0·2 in. The Hall voltage per gauss was about $1{\cdot}4 \times 10^{-5}$ for a current through the specimen of 18mA. Mason *et al.* (1953) used a plate 6 mm × 1·25 mm × 0·15 mm and a current of 1mA to achieve $2{\cdot}4 \times 10^{-5}$ volt per gauss. These small specimens permitted magnetic measurements to be made in small relays. Aigrain (1954) discusses the uses of germanium for Hall measurements of field and uses specimens of size similar to those of Mason *et al.* or smaller.

Recent investigations on the semiconducting compound indium antimonide show that it would be much more efficient than germanium (see table) but for the fact that at room temperature it is in the intrinsic conduction region. In consequence, the Hall coefficient and the resistivity are very dependent on temperature. Below about 150°K, this compound is an extrinsic semiconductor with a resistivity of about 1 ohm cm and a Hall coefficient of 3×10^{-5} (Tanenbaum and Maita 1953, Cunnell *et al.* 1953, Walker 1952, Weiss 1953). If the temperature dependence is not too severe a restriction, InSb might prove considerably better than germanium since it would be possible to pass a larger current through the specimen without the problem of overheating and the attendant thermal drifts. Schmeltzer† has found that it is possible to use currents up to 0·1 amp without temperature effects appearing in the InSb specimens and that the sensitivity seemed superior to that of germanium.

4.2. *Measurement Accuracy*

The equation quoted for the Hall voltage shows a linear relation between this voltage and the applied field. In general, this is only correct to a first order since there are four subsidiary effects which may contribute to deviation in the relation.

† Private communication: The author is indebted to Dr. Schmeltzer of CERN, Geneva, for drawing his attention to the use of InSb.

(i) Both the Hall coefficient and the specific resistivity of semiconductors are temperature dependent. These effects may be reduced to a minimum by the suitable choice of semiconductor material. Provided the semiconductor is not in the domain of intrinsic conductivity at the working temperature, the Hall coefficient is nearly independent of temperature. The upper limit for germanium is about 40°c and above this limit there is an ever-increasing temperature coefficient.

(ii) The specific resistivity is changed by the application of the magnetic field, i.e. the magneto-resistance effect. The ratio $\Delta\rho/\rho$ is approximately proportional to B^2 for small fields, but approaches a constant correction for large fields (Dunlap 1947). The magnitude of this change and the similar change mentioned in (i) can be negligible if the circuit is fed from a constant current supply.

(iii) The magneto-resistance effect also produces an effect on the Hall coefficient. In certain crystal orientations, the value of R may change by as much as 10% in germanium at 10^4 gauss. It is nevertheless possible to choose a specific orientation where R is field-independent and there are some simpler orientations with I and B along crystal axes where R changes by only 1% at 10^4 gauss (Mason *et al.* 1953).

(iv) Thermal e.m.f's generated by temperature differences at the contacts, and general contact variations, can contribute to small errors in the voltage. These show as annoying changes in the zero setting but may be eliminated in careful use. The elimination of drift due to thermal changes may be effected by using an alternating current on the probe (Donoghue and Eatherley 1951). This introduces a complication in that the contacts may be rectifying to some extent. Harmonics of the fundamental appear in the output voltage in consequence though these can be filtered out by the use of a frequency selective amplifier. In the region of low fields it is advantageous to use a phase sensitive detector as a null instrument. The calibration of the Hall instrument then depends on the fraction of the applied voltage required to balance the change produced in the specimen by the applied field.

4.3. *Summary*

Allowing for all these effects, it is still possible to construct a cheap, rugged instrument with a Hall probe. The measurement accuracy is approximately to $\pm 2\%$ up to about 10^4 gauss. Above this field, the variation of the Hall constant may be circumvented by the calibration of individual instruments. At low field strengths, however, the accuracy is limited only by the voltage measurement. Commercial Hall probes are available covering the range from 50 gauss to 20 kilogauss.

§ 5. Peaking Strips

5.1. *Introduction*

This magnetic field measuring device has been made possible by the advent of high-permeability low-coercivity ferromagnetic alloys, which, in fact, possess almost rectangular B–H loops. Such materials as Mo-permalloy and Mumetal have a maximum permeability in the region of 10^5 provided the specimen in use is not deformed or stressed subsequent to annealing.

The principle of operation of a peaking strip is shown by reference to figure 11. A strip of high permeability alloy is placed in a uniform magnetic field which

increases linearly with time (figure 11 (*a*)) and its length along the direction of the field. Figure 11 (*b*) shows the (*B*, *H*) loop for a small section of the strip

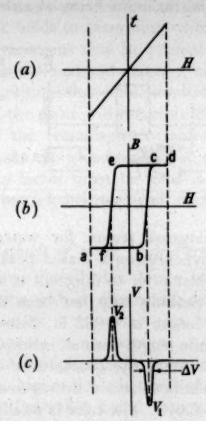


Figure 11. Demonstrating the production of voltage pulses on traversing the (*B*, *H*) loop of a high permeability material. See text for description.

at the centre, where a search coil of N turns is wound close to the strip. The voltage induced in this coil at any instant will be given by

$$V = NA(dB/dt) \times 10^{-8} \text{ volts}$$

where B is the flux density in the strip at any instant in gauss, and A is the cross-sectional area of the strip in cm^2. As the value of H changes from point a to b (figure 11 (*b*)), the strip remains saturated and there is no flux change through the coil. At b, the flux density begins to alter and passes rapidly through zero to a value equal to that at b but with opposite sign, i.e. to point c. The passage through zero B marks the maximum dB/dt and hence maximum voltage output V_1. The final stage of traversal to point d will result in no output voltage. The reverse procedure takes the strip through the path d e f a on the (*B*, *H*) loop, giving a voltage pulse V_2, figure 11 (*c*). Although each of these voltage pulses appears at a time when the field is not zero, this presents no serious problem in measurements. In a field which is oscillating about zero, the voltage pattern will be symmetrical if displayed on an oscilloscope whose spot is swept by either a magnetic or an electric field which is in phase with the oscillating magnetic field applied to the peaking strip. The centre of symmetry of the pattern is taken as reference and is made to correspond with the undeflected position of the spot (figure 12). In a field which is rising, the displacement from the zero field condition can be absorbed into the calibration of the device provided there is no alteration in the rate of change of field at the instant of measurement.

To change the instant at which the voltage pulse appears, a biasing field B_0 can be supplied which displaces the axis of B in figure 11 by B_0. In the case of

the oscillating field, the pattern of figure 11 (*c*) will be made to move across the oscilloscope trace. The limitation in providing this field is to be found in the power dissipation in the coils, the limit being at several hundred gauss for air-

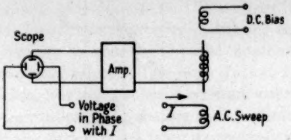


Figure 12. Schematic circuit for the production of signals from a peaking strip.

cooled coils and in the kilogauss region for water- or oil-cooled coils. In Birmingham on the proton synchrotron a biased peaking strip is used to define an instant in the field cycle just prior to injection at about 200 gauss.

A very reliable form of peaking strip has been described by Kelly (1951). This consists of a two-inch length of 0·002 in. diameter Mo-permalloy, sealed inside a quartz tube of the same length with an inside diameter of 0·01 in. and an outside diameter of 0·05 in. With the quartz tube open at one end, the wire is inserted and annealed in an atmosphere of hydrogen at the temperature specified by the manufacturers of the alloy. The tube is finally sealed off and the search coil wound directly on to the outside of the quartz tube. With a search coil of 2000 turns and with dB/dt about 3000 gauss sec^{-1} an output pulse of 100 millivolts is obtained, the rise time being about 70 μsec.

5.2. *Measurement Accuracy*

Kelly (1951) has investigated the factors which affect the performance of a peaking strip and has considered the following:

(i) The existence of free polarity on the strip due to its finite length decreases the effective permeability. The demagnetizing factor is such that length to diameter ratios for cylinders of this alloy should be in the region of 1000, at least, for best results. He therefore chose to use a wire of 2 in. length and 0·002 in. diameter.

(ii) The effective permeability is a maximum at the centre of the strip and falls off on each side. Since the coil is of finite length, the mean flux change through the coil is not as high as the maximum possible. Generally, however, it is possible to wind the coil at the centre of the strip with very fine wire, making the winding perhaps 5% of the overall length.

(iii) Gradients in the applied fields can cause various segments of the strip to go through maximum permeability at different times. This broadens the voltage pulse in time and can reduce its magnitude.

(iv) Eddy currents induced in the strip by the changing field produce an effect analogous to the skin depth in conductors. Flux is prevented from penetrating fully into the strip with a consequent reduction in the total flux changing through the coil. Kelly indicates that for dB/dt equal to 10^4 gauss sec^{-1} or less, the 1/2 power width of the pulse from the type of peaking strip he describes is about 0·2 gauss. Above this value of dB/dt the pulse tends to broaden from eddy current effects.

It has been found that by displaying the pulse on an oscilloscope,

$\pm 0{\cdot}02$ gauss setting accuracy can be achieved. The application of these peaking strips to field measurement is dealt with by Giordano *et al.* (1953). Bosley *et al.* (1947), Clark (1946) and Adams *et al.* (1947) describe a method of utilizing peaking strips in betatron magnetic fields to determine when the field is passing through zero and what azimuthal variations may be present at that time. However, the presence of other ferromagnetic material makes it necessary to apply a correction factor since under these conditions flux collection to the strip is not so pronounced. The time of appearance of the pulse will therefore depend on the relative position of the peaking strip and the ferromagnetic material. Furthermore, if a bias field is used, for example in the measurement of the remanent field of a large magnet, another correction factor must be used due to the proximity of ferromagnetic material which changes the magnetic reluctance between the ends of the coil.

5.3. *Other Methods of using High Permeability Material*

Adams *et al.* (1950) have described a device in which a ferromagnetic strip is excited to partial saturation with an alternating current. The second harmonic content of the exciting current is measured and, when the strip is biased by a small external field, it appears in an amount proportional to this field in the region up to 10 gauss. Although this is essentially a low field detector, the method could be applied to higher fields in much the same way as a peaking strip.

Gregg (1947) uses a Permalloy-cored transformer in a similar way for measuring milligauss. Although not directly applicable here, such methods were developed for use in the detection of small magnetic fields (e.g. from magnetic mines) but could be adapted for use in a way similar to the peaking strips. For a very comprehensive set of references the reader is referred to the paper of Kühne (1952).

5.4. *Summary*

The peaking strip method of obtaining an electrical impulse at a time related to a given magnetic field provides an inexpensive and simple way of measuring both steady and dynamic fields over the range from zero to a few kilogauss accurate to about 0·02 gauss. The high permeability alloy is available commercially and the amount of associated equipment is small, well within the range of the smallest laboratory.

§ 6. The Generator Principle

6.1. *Rotating Coil Fluxmeters*

When a coil is moved in a magnetic field in such a way that the total flux linking it changes, a voltage is induced in the coil proportional to the rate of change of flux and the turns–area product of the coil itself. This is, of course, only a restatement of the well-known principle of the generator but it provides a very simple and accurate means of field determination, though in general a calibration should be made by reference to some absolute method for maximum precision. If the coil movement is a rotation in a steady field, the alternating output voltage is used for determination of the magnitude of the field. If the coil is moved back and forth in a simple harmonic motion, on a line normal to its axis, the voltage output will be proportional to the frequency and to the gradient of the field component parallel to the coil axis.

(*a*) *A one-coil system.*

The coil is rotated about an axis perpendicular to the axis of the coil at a constant angular velocity ω. The alternating output voltage appearing at the terminals will be $V=10^{-8}\,NA\omega B_n$ volts where N=number of coil turns, A=mean area of one turn in cm², ω=angular velocity, B_n=component of field perpendicular to axis of rotation in gauss. The alternating voltage may be collected by slip-rings, amplified, rectified and supplied to a d.c. voltmeter to give a reading proportional to B_n provided ω is maintained at a constant value to the required accuracy (Jürgens 1953).

Alternatively, the output may be taken off by means of a commutator and brushes to a direct reading millivoltmeter (Kohaut 1937, Cole 1938, Peters 1950). Cole states that an accuracy to 0·1% is obtainable with the use of a potentiometer for measurement of the voltage output.

Cork, Shreffler and Shull (1947) have constructed the rotating coil as a set of coils wound as in a lap-wound d.c. generator with 24 silver segments as commutator and silver brushes to collect the voltage. Konakhovitch *et al.* (1949) have a similar coil with 48 segments on the commutator and have obtained relative accuracies to 0·02%.

In this same category is the Faraday disc or homopolar type of generator. A conducting disc rotating with its axis parallel to a magnetic field develops a voltage between an inner and the outside radius, proportional to the angular velocity and the magnetic field. Although the voltage levels are small, it is possible to calculate the voltage output if the two radii are known accurately (Caro, private communication).

In each of these d.c. cases, there is the problem of preventing errors creeping in from thermo-e.m.f's and contact noise. This can be overcome in the rotating coil case provided it is permitted to increase the turns–area product to give voltages in the range of volts. Commercial instruments are available with a low range of a few hundred gauss full scale on a comparatively long scale meter. The instrument is robust and, by using a coil of very small overall dimensions, accessibility to narrow magnet gaps is claimed as one of its great virtues. Where stable frequency mains are available, the drive may come from a small synchronous mains voltage motor.

To overcome the problem of contact noise and thermal voltages, Klemperer and Miller (1939) constructed a coil which was given an oscillation by 180° about an axis in its plane. This was done by using a reciprocating mechanism to drive a pair of gear wheels on the last one of which the coil was mounted. The whole mechanism was driven by a synchronous motor. In this case, the coil leads are taken away directly, the alternating voltage output being rectified and supplied to a meter. They achieved a sensitivity of 3 μA per gauss and were able to study the axial and radial field configurations of electron lenses. A further reference to this work is given by Goddard and Klemperer (1944).

(*b*) *A two-coil system.*

In the single coil system, there is always the problem in calibrating the instrument of maintaining a constant frequency supply to the driving motor. If, however, a second coil of known turns–area product rotates at the same angular frequency and phase in a known magnetic field, its output voltage may be balanced

against the voltage from the coil in the unknown field and the calibration no longer depends on the frequency of rotation of the coils.

The standard magnetic field may be provided simply in two ways. Hedgran (1952) and Langer and Scott (1950) have used a pair of Helmholtz coils so that the comparison comes to a measurement of current through these coils. Caldecourt (1950), Ploch (1949), Katz *et al.* (1950), van der Walt (1953), and Lamb and Retherford (1951) have used a permanent magnet (calibrated by some other standard means) to achieve a stable reference field, the output voltages being balanced through a potentiometer. Parkinson *et al.* (1947) use an auxiliary magnet fed in shunt from their main magnet to balance first order variations.

The precision with which the outputs can be balanced and the accuracy of the current or standard field measurement determines the overall accuracy of the method. Suppose the voltages from the two coils are $(v+\Delta v)\cos\omega t$ and $v\cos(\omega t+\pi+\delta)$ where the balance in phase is made mechanically or otherwise to the accuracy of δ. The output after balancing is then

$$\begin{aligned}\Delta v_0 &= (v+\Delta v)\cos\omega t + v\cos(\omega t+\pi+\delta)\\ &= v\,(\sin^2\delta+\Delta v/v)^{1/2} \text{ neglecting higher order terms.}\end{aligned}$$

Hedgran (1952) shows by plotting $\Delta v_0/v\sin\delta$ against $\Delta v/v\sin\delta$ that the minimum may be located to a region $\Delta v \simeq v\sin\delta$. In his case, the attempt to achieve an accuracy to 2 parts in 10^4 meant that δ had to be less than $0{\cdot}003°$. In this adjustment of phase opposition the overall torsional vibration between the coils during the measurement runs contributes to the value of δ as does the static mechanical accuracy of alignment. Hedgran discusses this problem in detail and reference to this paper is well advised for constructional details. In any case, the adjustment of phase opposition to the specified phase accuracy is difficult mechanically and the use of a weak cross-field B_y applied perpendicular to the reference field B_z gives the final critical adjustment necessary. The application of this field does not affect the accuracy of the main field since the correction field is usually less than $0{\cdot}2\%$ of the main field. The sensitivity achieved was about 10^{-3} gauss in the range up to about 200 gauss. Ploch (1949) quotes accuracies for his more standard instrument of 50% at $\frac{1}{2}$ gauss, 1% at 50 gauss and $\frac{1}{4}\%$ at higher fields. Van der Walt (1953) has used this system for magnet stabilization as well as measurement accurate to $0{\cdot}01\%$. Katz *et al.* (1950) obtained an accuracy to better than $0{\cdot}1\%$ with a similar arrangement. The work on the fine structure of the hydrogen atom has required the measurement of magnetic field to considerable precision. Lamb and Retherford (1951) have constructed a rotating coil fluxmeter using a permanent magnet reference field and they discuss the problems of obtaining relative accuracies to 1 in 5×10^4. A serious problem subsequently presented itself in that the calibration constant of the system is temperature sensitive. This is discussed by Dayhoff *et al.* (1953) in a companion paper.

In general, the unknown field is not homogeneous and the pickup coil takes a mean value over its area. The voltage will nevertheless be very nearly sinusoidal. Ploch discusses this problem and shows how the gradient terms may be determined to correct the measured value to give the true value at a point. Probably a more straightforward method of obtaining the field at a point would involve the use of a fluxball coil or a simple cylindrical coil of the correct dimensions (see § 7.2).

If, instead of using the two similar coils separately, they were both used in the unknown field separated by some known distance, it would be possible to measure the field gradient by the voltage remaining after they had been connected in opposition. Here it would be of greater importance to consider the use of coils of the correct shapes to give point measurements in each case. Grivet and Regenstreif (1953) quote the sensitivity of such a device used in their laboratory as 3·5 mv r.m.s. for a field gradient of 1 gauss mm^{-1}.

6.2. *Vibrating Coil Fluxmeters*

It has become increasingly important in electron lens and nuclear machine design to measure with some accuracy the field gradient as distinct from the magnitude of the field itself. A coil which vibrates back and forth in its own plane and with its plane normal to the magnetic field will record the flux change through it as an alternating voltage proportional to the field gradient, the frequency of oscillation and the amplitude. It is therefore necessary to provide, as well as a stable frequency power supply to drive the electrodynamic vibrator, a system for controlling the amplitude of the vibration accurately. Unfortunately, large magnets commonly have quite large leakage fields near them so that it may be necessary to move the vibrator some distance away unless magnetic screening can be employed. This problem is discussed with others by Grivet and Regenstreif (1953).

6.3. *Summary*

The rotary generator type of fluxmeter presents an excellent solution to the problem of precise field measurement in the region of hundreds of gauss where the limiting sensitivity to changes has been found as low as 10^{-3} gauss. In its simplest form, it is a portable device which may be run off the a.c. mains giving reproducibility mainly limited by the frequency stability of the supply. The twin rotating coils and the vibrator provide a fairly easy method of field gradient measurement in both small and large magnets.

§ 7. Fluxmeter Methods

7.1. *Mechanical Fluxmeters*

Of all the diverse methods of field measurement available at the present time, the fluxmeter of the Grassot type and the ballistic galvanometer still remain reliable and quite usual field measuring instruments in practice. The Grassot type fluxmeter as commercially available to-day is a much more elaborate instrument than its predecessors since it is common to find that drift-correction devices and zero-setting circuits are built in with the instrument.

In a fluxmeter, the voltage generated by the change of magnetic flux through a search coil is countered by a back-voltage generated by the meter movement. Provided the inertia, electrical resistance, suspension torque and air damping give negligible effects on the motion, these two counter voltages will be equal and integration of the equations of motion shows that angular displacement is directly proportional to the difference between the final flux and the initial flux through the search coil. The small effects mentioned determine the performance of the fluxmeter as an instrument which can give this proportionality in all parts of the scale. Calibration is readily achieved by the use of a standard mutual

inductance, the secondary of which is directly connected to the fluxmeter and whose primary is subjected to a known current change (Nettleton and Sugden 1939, Cook and Mather 1947).

A careful and systematic treatment of the problems of drift, rapidity of response, sensitivity and torque compensation is given by Cook and Mather (1947) in their study of a fluxmeter for magnetic measurements on the 164-in. Columbia synchrocyclotron magnet. The fluxmeter used was found to be very linear and, after adjustments to the torque compensator and insertion of a small voltage to counter thermally induced voltages, the drift was reduced to less than 0·01 mm per second in all parts of the scale. Reading accuracy was generally to about 0·1%. A similar treatment is given by Dunn (1939) in a discussion of the Grassot fluxmeter as a quantity meter. Spilsbury and Webb (1945) deal with problems in the calibration of a fluxmeter and give suggestions about avoiding errors.

If a ballistic galvanometer is used large errors can arise from lack of reproducibility in the time taken to move the search coil and from variations in the ratio of the time taken to move the coil to the period of the galvanometer. If the coil is rotated through 180° to measure the magnitude of the field, constancy of the angle of rotation is of prime importance (this applies to Grassot fluxmeters as well). These points can be met if the coil is moved mechanically and the period of the ballistic galvanometer is long. Inozemtcev and Latishev (1949) find that it is possible to achieve 0·1% reading accuracy in this way.

It is possible to devise methods using a combination of search coil and mutual inductance which permit the fluxmeter or ballistic galvanometer to be used as a null instrument. Nettleton and Sugden (1939) used a switch connected to the search coil support. The contact was broken on moving the coil, cutting off the current flow through the mutual inductance primary. Adjustment of the current to give a null immediately gave a proportional measure of the field strength. A null method giving a relative answer (Foss, private communication) involves the use of two coils, a fixed distance apart, which are made to oscillate through 180°. The axis of oscillation is through the midplanes of the coils and perpendicular to the coil axes. The voltage outputs of these coils are taken to potentiometers, the outputs of which oppose each other through a galvanometer. Since it is not possible to align the coils exactly, a third compensation coil is fixed to the oscillating shaft with its axis perpendicular to the axis of oscillation and the main coil axes. This is connected into the circuit shown in figure 13. In normal

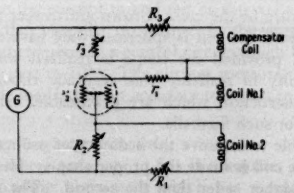


Figure 13. A null method circuit for relative field measurement in which the coils are made to oscillate through 180° (from notes by J. Mates, Carnegie Institute of Technology: private communication by M. Foss).

operation r_2 is set at 10^4 ohms, the coils placed so that they are in equal fields and the other resistances adjusted for minimum galvanometer deflection. Any final unbalance is due to coil misalignment and is eliminated by adjusting the compensating resistances r_3 and R_3. After a good balance is achieved, the coil support is placed so that coil 1 is in some position of importance, for example, the magnetic centre of the field and on oscillating the coils a balance is reached by using r_2. The reading on the resistance box r_2 taken relative to 10^4 ohms will give the field at coil 2 relative to that at coil 1. Other points may be found in two ways. Transferring the coil 1 to the position first held by coil 2 and repeating the measurement gives a series of readings spaced by the distance between the coils. To obtain the intermediate readings, a guess must be made for the value of the change between the standard position and the intermediate position. Coil 1 is placed in this position and the relative measure made once more. The use of these two methods enabled Mates to reach relative accuracies to 0·04%. Techniques which are similar in principle have been used on many of the large magnets associated with cyclotrons. Relative measurements can be made by balancing the outputs from two coils, via a potentiometer. Such techniques have been used on betatrons to determine radial field variations (Finlay *et al.* 1950).

7.2. *Fluxmeter Coils*

It has been usual to wind coils on a cylindrical former in such a way that it is practicable to calculate the effective area. If this is not sufficiently reliable the coil may be calibrated in a Helmholtz coil system whose current–field constant is known by geometry or has been measured by some standard technique such as nuclear resonance. Coil calibration has been discussed by Nettleton and Sugden (1939) and by Mikhail and Yousef (1945). Associated with the use of these coils comes the problem of deciding what the measurement means in terms of the average value of an inhomogeneous field, given by the fluxmeter reading. It can be shown that if the value of the field at the centre of the coil is B_z, the total flux Φ through the coil of area A and turns N is given by

$$\Phi = NAB_z\left(1+\frac{l_x^2}{B_z}\frac{d^2B_z}{dx^2}+\frac{l_y^2}{B_z}\frac{d^2B_z}{dy^2}+\frac{l_z^2}{B_z}\frac{d^2B_z}{dz^2}+\text{higher order terms}\right) \text{ where } l_x, l_y, l_z$$

are proportional to the coil dimensions in the x, y, z directions. Since the coil is being used in a source-free region, we have $\nabla^2 B_z = 0$. Thus, if $l_x = l_y = l_z$, the second-order terms vanish. This ensures that the mean value of the field measured would refer to the centre of the coil. Brown and Sweer (1945) and Williamson (1947) have shown that if the coil is spherical (they have named it a 'fluxball'), all corrections vanish provided the shape is realized with sufficient accuracy. There is some difficulty in realizing this accuracy though both authors have devised means of construction which are practicable. Brown and Sweer give precise winding data for such fluxballs.

It should be possible to improve the accuracy of ordinary measurements with a cylindrical coil if the coil is made the proper shape. Here it will be important to consider terms of higher order than the second. The obvious constructional advantages of the cylindrical winding have led Verster (1949) and Herzog and Tischler (1953) to consider the use of single and multilayer cylindrical coils respectively. Verster used a coil with a length/diameter $= \sqrt{3}/2 = 0{\cdot}866$ to achieve

elimination of the second order terms. Herzog and Tischler calculated the shape for a multilayer coil whose inner diameter was zero. The experimental determination of the best coil shape gave a length/outer diameter ratio of 0·72 assuming an inner diameter of half the outer diameter.

Verster and Henneberke (private communication) have shown that it is possible to achieve a compensation of the second order terms and a partial compensation of the fourth order terms for multilayer coils. They give a graph showing the values of a/r_2 for values of r_1/r_2. Using the correct dimensions from this graph, they show that

$$\Phi=\frac{\pi N}{3}\frac{r_2^3-r_1^3}{r_2-r_1}B_z\left\{1-\frac{B^{iv}}{B_z}C r_2^4+\ldots\ldots\right\}$$

where $2a$=coil length, r_1 and r_2 are inner and outer radii of coil, and C is a constant determined from a corresponding graph again plotted against r_1/r_2. To design a coil of the best shape making $C\leqslant-2\times10^{-3}$, r_1/r_2 should be less than 0·3, for which condition the value of a/r_2 should be 0·67. The advantages of using such coils for fluxmeter measurements are obvious but it should be noted that they would also improve the usefulness of rotating coil fluxmeters.

7.3. *Magnetometers*

(*a*) *A null-reading instrument.*

Lauritsen and Lauritsen (1948) have devised a simple and reliable magnetometer. The principle of the instrument is that the current required to hold a coil in a specified orientation against a constant torque is measured. Strictly speaking, this instrument should be classified as a torsional current balance but because of the name given to it, it is inserted here. The indicator is first set to zero with the field on and no current through the coil by rotation of the torsion head. A selected fixed angle of rotation is set and the coil current adjusted to restore the indicator to zero. The fluxmeter described in this paper required 15 mA to restore equilibrium for about 7500 gauss. Over a period of a year the calibration changed by less than 0·2%. The restoring torque was supplied by a phosphor bronze spring which in the end did prove unsatisfactory, due to hysteresis and drift limiting the accuracy of measurement. Milne (1951) later improved the instrument by using a quartz fibre. Comparison figures for quartz and phosphor bronze are given as 0·05% and 0·4% hysteresis, and $1{\cdot}3\times10^{-4}$ deg^{-1} and 6×10^{-4} deg^{-1} rigidity temperature coefficient respectively. Phosphor bronze springs are used to take the current in and out of the coil and to keep the quartz fibre in tension but they supply no torque to the system. Milne found it possible to use the magnetometer in a control circuit which held the magnetic field steady to better than 1 part in 10^4. The current measurement enabled a relative field measurement to be made with an accuracy to 1 part in 10^3.

(*b*) *Magnetometers using magnetic material.*

For measurements where accuracy to 5% is adequate, it is possible to obtain a commercial direct reading instrument which has a small permanent magnet mounted on a pivoted shaft. The magnet is made of a high coercivity alloy, the coercivity being about 6000 oersteds. The direction of magnetization is across the diameter of the shaft and the instrument pointer is adjusted parallel

to this direction. The shaft is controlled by a coiled spring and this holds the pointer at zero position on the scale in zero field. In use the instrument is held so that the direction of the field is at right angles to the instrument pointer, a position which is found by orienting for maximum reading. The direct reading is taken from the scale, the polarity being indicated by the direction of movement of the pointer. By orienting the meter to give a zero reading, the pointer indicates the direction of the field. Different instruments are used for different orders of field up to a maximum of 3000 gauss. Field values in excess of this may upset the permanent magnet and change the calibration of the instrument.

Briggs and Mitchell (1949) have designed an instrument which relies on the torque imposed on a strip of soft magnetic material in the presence of a magnetic field. To avoid calibration problems the permeable material should have a low hysteresis and remanence. This suggested the use of dust iron mounted on a non-magnetic base. In fact, they used a pair of concentric cylindrical vanes of aluminium covered with a very thin layer of iron powder, the whole being supported on a pivoted movement with coil springs for supplying the torque. At 1000 gauss, the torque is sufficiently large to permit the use of robust springs and ball-ended pivots. With only 1 mm^3 of material of permeability 10 in a field of 1500 gauss, a torque of about 60 g cm was obtained. The instrument was found to be reliable to 2% and with strong springs could be made to work up to 7500 gauss, at least. Another such meter was designed for use in a gap of the order of 0·05 in.

For fields of the order of 100 gauss, moving-iron a.c. ammeter movements (no external connections) may be calibrated for use as fluxmeters. Here the problem of hysteresis and remanence of the movement is well understood so for rough work to 5% this gives a simple method of measurement. The use of direct current through the meter winding also permits the meter to be used as a null instrument.

7.4. *Electronic Fluxmeters*

(*a*) *Steady field measurements.*

The main function of the Grassot type of fluxmeter is to integrate the voltage generated when a search coil is moved from one position to another where a different total flux links it. There are many ways of doing this electronically and several methods have been described which rely basically on the integrating properties of a simple RC circuit (Harris 1931, Gregg 1949).

In another method, Edgar (1937) showed how a galvanometer could be used as an instrument giving an approximate null, by opposing the voltage due to the flux change in the search coil with one obtained by an equivalent flux change through the secondary of a mutual inductance. He accomplished this by splitting the light beam from the galvanometer between two photocells connected in a bridge circuit and applying any voltage unbalance, resulting from a deflection of the galvanometer, to the grid of an amplifier valve whose anode current passes through the primary of the mutual inductance. The change of the primary current in operation is a measure of the flux change through the search coil. The circuit behaves as if it had an infinite time constant so that a meter reading may be taken easily. Dicke (1948) and Green *et al.* (1953) have adapted the circuit to be used with a double photocell. Cioffi (1950) has developed the circuit to be used with a recorder to give an instrument of high accuracy and sensitivity. This

device has a double circuit which, with the Speedomax recorder (Williams 1938), enables (B, H) loops to be plotted directly with accuracy on to paper, the scale adjustment being easily made to suit requirements. Green *et al.* (1953) used a self-balancing potentiometer to record the required answer.

(*b*) *Dynamic field measurements.*

In the previous section it was assumed that the flux linkage changed from one value to another as the search coil was moved in space. When the flux linkage with the search coil is changing in time and the value of the flux density at any instant during this time is required, it is possible to use an integrating circuit which consists of a d.c. amplifier with capacitive feedback. This is the so-called Miller circuit which is described by Chance (1949). An excellent description of an electronic integrator together with a theoretical analysis of simple circuits as well as the circuit used is given by Ekelof *et al.* (1951). The main advantage of these electronic devices lies in the ability to obtain a longer integrating time constant than is possible with a simple RC circuit while maintaining the sensitivity, or conversely, the same integrating time constant as can be achieved with RC circuits with an enhanced sensitivity. Furthermore, the output of integrating amplifiers may be displayed on an oscilloscope enabling the study of transient phenomena to be made. Ekelof considers its use for measuring the magnetic flux through the yokes of relays under operating conditions. Green *et al.* (1953) describe a similar circuit for measuring the magnetic field of the Brookhaven Cosmotron magnet and give a discussion of the errors involved in the measurement.

Another method employing the basic Miller circuit is described by Fuller and Hibbard (1954). The linear fall of anode voltage in the Miller circuit is made to occur between two well-defined limits. At the lower limit, a commutating circuit returns this Miller unit to the upper limit of anode voltage and starts another similar unit on its linear fall of anode voltage. At the end of this run-down, the first unit takes over again and so on. This oscillating circuit operates with a frequency which is defined by the input voltage, since this voltage determines the rate of fall of anode voltage. The relation between frequency of operation and the input voltage can be made linear to about 0·1% over a given region of input voltage. A set of scalers counts the number of oscillations, which summation gives a figure proportional to the integral of the applied voltage. A system for reading the scalers at any number of specified instants in time is described. The system has been applied to the measurement of the rising magnetic field of the Birmingham proton synchrotron magnet at approximately 10 millisecond intervals during the one-second rise from remanent field to 12 000 gauss. The accuracy achieved is to about 0·1% and has been checked by assessing how accurately the accelerated particles stay in the centre of the vacuum box at all times during the cycle of acceleration. A modified system of this nature is under construction for the measurement of dynamic fields in the CERN alternating-gradient synchrotron at Geneva.

7.5. *Summary*

In the measurement of magnetic field, Grassot fluxmeters and ballistic galvanometers still play an important role and it is possible to achieve a measurement accuracy to 0·1% if due care is exercised in setting up the instrument and

circuit well. Fluxmeter coils can now be designed to give point measurement in an inhomogeneous field, the coils being either spherical or of special cylindrical shape.

Magnetometers of a null-reading type, which have given reliable service over a considerable period, are in use. Relative measurement accuracies to 0·1% have been obtained and the instrument has been built into a magnetic field control circuit with good effect. Magnetometers using ferromagnetic material are now commercially available and give the possibility of obtaining readings quickly to a few per cent.

Electronic integrators have been devised and for dynamic field measurement these are superior to any other device. They have been used for displaying the rapid changes in field which occur on the excitation of a relay as well as for measurements on the magnets of particle accelerators where the magnetic field rises to its peak value in about one second. One method, at least, shows that measurement accuracy to the order of 0·1% can be reached. In all these integrating methods, it should be noted that the use of a pair of coils enables the gradient of the field to be measured.

8. Forces on Current-Carrying Conductors

8.1. *Current Balances*

When a conductor carrying current is placed in a magnetic field, it is subjected to a force which is proportional and at right angles to both the current and the magnetic field. This principle when applied gives one of the most fundamental and accurate methods of measuring magnetic fields. When the magnetic field is horizontal, the balance method of Cotton and DuPouy (1932) may be used. A rectangular coil is hung from a balance so that its plane is normal to the field and only the lower side of the coil is wholly in the magnetic field. When a known current is passed through the coil, it is the force on this lower horizontal section which is measured with the balance. Since the effective length of the horizontal section can be determined very accurately, as can the current through it, the measurement of the force knowing the acceleration due to gravity is all that is needed to complete the field determination. Several other people have used the method at various field strengths (Briggs and Harper 1936, Briggs 1936, Scott 1934, Rogers 1936, 1937, Slätis 1953, Thomas *et al.* 1950). At fields above about 5000 gauss it seems possible to achieve an absolute accuracy to 1 part in 10^5 though this would require considerable patience in eliminating all sources of error (see Thomas *et al.* 1950). This latter paper discusses all the pertinent problems in some detail besides giving the basic theory of the measurement.

It is of course possible to use the current balance idea for measurements requiring a much lower order of accuracy. Chang and Rosenblum (1945) have adapted the method for use with the vertical field of a cyclotron and show that a sensitivity of 1 part in 10^4 may be achieved. The absolute accuracy was to about 0·2% without very elaborate precautions being taken. Another form of instrument is also described by which small changes in a magnetic field are measured. This relies on having a non-linear region for the gross variation of the field from zero up to the working region at which point the sensitivity has markedly increased. This is achieved with a rectangular coil whose plane is initially normal to the

field but is nearly parallel to the field in the working position. With this device, they were able to detect a change of 2 gauss in 12 000 gauss easily. The so-called magnetometer of Lauritsen and Lauritsen, described in § 7.3 (*a*) is a somewhat similar instrument which, because of the name given it by the authors, has been placed with magnetometers.

Kaufmann (1938) describes another form of current balance which consists of a single layer solenoid whose axis is along the magnetic field, one end being in the strong field B_0 while the other is in the weak field B_w outside the magnet. It is shown that the force is given by

$$F=\frac{niA}{10}\ (B_0-B_w)\ \text{dynes}$$

where n=turns per cm of coil, i=current in amps, and A=the cross-sectional area of the coil in cm^2.

The value of the force is independent of the pitch of the coil. Relative measurements will have a very high accuracy while absolute measurement to better than 0·1% can be made. A variation of this method is used by van Ments and Le Poole (1947) for measurements on electron lenses. They constructed a long thin solenoid in which the windings reversed at the midpoint of its length. The solenoid is long enough to project from each end of the electron lens under investigation so that the extremities are in almost zero field. It can be shown that the horizontal force acting on the solenoid which is suspended horizontally by thin wires causes a displacement of the midpoint. This displacement is a measure of the field strength at the position of the midpoint in the field. In general, the system is used to determine the field strength along the axis of an axially symmetric system where it is possible to estimate the value of the derivatives of the field approximately from the measurements. These estimated values are used to compute corrections to give the true field strength at the midpoint of the windings. If a solenoid of very small radius is used these corrections become almost negligible. Van Dorsten and Franken (1953) adapted this method for use with a solenoid carrying alternating current. In this case the coil is suspended with its axis vertical in such a way that the vibrations resulting from the action of the field on the current are transmitted to an ordinary gramophone piezo-crystal pickup. The output is amplified so that relative measurements may be made accurately.

Birebent (1952) has shown how a piezocrystal may be used as a mechanical detector of the force exerted on a wire loop carrying alternating current. The 'hairpin' has an active element 2·5 mm in length. The mechanical system is designed to resonate at the frequency of the alternating current to enhance the sensitivity. The first method of measurement involved adjusting the current to give a constant voltage output from the crystal. The product of field and current is then a constant which may be determined in a known field. Experience showed that the response to various voltages was in fact linear and a second method involved holding the current constant and measuring the output of the crystal. This was a more rapid way of using the instrument. With the amplifier used, the range of measurement covered was 30 to 20 000 gauss. At low fields, trouble was experienced in attempting to eliminate sources of external noise which were recorded by the crystal. This also applies to the method of van Dorsten and Franken.

8.2. *Particle Trajectories: the Floating Wire Method*

Suppose a perfectly flexible wire is attached to a support which allows it to assume any direction it desires, and is stretched over a pulley whose support is also free to pivot about an axis perpendicular to the wire. The tension T in the wire is supplied by a weight hanging on the free end of the wire. A current is passed along the wire which is placed in a magnetic field B. The wire takes up a curvature of radius ρ cm, which, at any point along it, is related to the magnetic field and the current by $10T/i = B\rho$, where T is in dynes, i in amps and B in gauss. It has been shown by Loeb (1947) that there is an analogy between the behaviour of this wire with a given tension and current, and that of a beam of charged particles of given momentum travelling in the same field. In fact, if we consider a beam of charged particles with momentum p (in units of ev/c) in a magnetic field B gauss their path will have a radius of curvature ρ given by $p = 300B\rho$. Comparing the two formulae, we have $p/300 \equiv 10T/i$ in the units specified. Loeb discusses the application of the analogy to plotting particle trajectories through magnetic lenses and describes the details of apparatus for given particle momenta. It seems that there is little further in the published literature on this subject though the author is aware of a considerable number of laboratories where extensive use is made of the technique for determining the paths of particles of given momentum from accelerators.

There is, however, an unpublished report by Cranberg (1951) on magnet calibration using this technique. He describes how it is possible to define the path of a beam of ions, set up a wire to follow the same path, then measure the ratio of wire-tension to current, while the magnetic field is held steady at the value at which the path of the ion beam was measured. The main sources of error in applying the analogy come from the assumption that the force on the current due to the presence of the magnetic field is the only interaction on the wire. He finds that there are three other forces to be considered: (*a*) the extra forces on the current resulting from the self inductance of the loop in which the current is several orders of magnitude larger than the ion or electron beam current, (*b*) elastic forces due to the stiffness of the wire, and (*c*) the effect of gravity on the wire itself. The error due to (*a*) is shown to produce an error in the tension of the order of 1 mg if the wire current is 1 ampere, the radius of the wire 0·1 mm, the pole separation about 4 cm, and the radius of the path about 1 metre. To minimize the effect of stiffness, the wires were annealed by passing a heavy current while they were hanging on the prescribed path, thus removing all the irregularities. Cranberg points out that the most troublesome effect is due to gravity. A plot of wire-tension versus wire-current yielded a straight line which gave invariably a positive intercept on the tension axis, which became greater the heavier the wire used. A detailed calculation was made and it was shown that, including gravity, the slope of the tension–current curve still gives the particle momentum. The calculation also gives the value of the intercept on the tension axis, and the true wire trajectory. Practical details given in this paper are comprehensive but beyond the scope of this Report. Results are given for runs with aluminium wire of 0·0082 inch and 0·0055 inch diameter with tensions between 20 and 100 g. For good life, the current in the latter never exceeded about 1·5 A in operation though an annealing current of 1·8 A was used for 1 minute. It is shown that an r.m.s. error in the determination of the slope

of the tension–current curve could be as low as a few hundredths of one per cent so that the momentum of the particles can be determined with the same precision, with the apparatus described.

8.3. *Median Plane and Magnetic Centre Determination*

The University of Chicago (1949) and Mates (1951) use a wire loop carrying current, supported at one point by a hydrometer of given force constant to find the magnetic median plane at the centre of a cyclotron magnet. To determine the median plane position, the level of the liquid supporting the hydrometer is altered until the vertical position of the point of suspension does not alter when the current is varied. If for a given current the loop is above the magnetic median plane, an increase in the current would cause the loop to rise, and vice versa. When the height remains unaltered, the point of suspension is at the median plane. In estimating the magnetic centre, the loop is allowed to move horizontally in the gap of the magnet, the field being vertical, until a steady position is reached. Location of the position of the loop gives an indication of the magnetic centre. However, the restoring torque is extremely low and it was found necessary to adopt a technique of successive approximations. With the support at 0° azimuth, the apparent magnetic centre was determined for several wire currents. The support was then moved to the 90° azimuth and a further set of readings taken. A plot of these results gave two lines at right angles along which the apparent centres lay for the different currents. The intersection of these lines gave the true centre.

8.4. *Summary*

Current balances provide an accurate and absolute measurement of large magnetic fields to the order of one part in 10^5. At lower accuracies, instruments have been designed to give excellent reliability. By using piezocrystals as force detectors, devices have been constructed for relative field measurements. These cover the range from ten to ten thousand gauss with reasonable ease.

By analogy, it is possible to use a stretched current-carrying wire for several purposes which are not essentially field measurements. Particle trajectories in a magnetic field are readily found and a straightforward determination of the magnetic median plane and magnetic centre of a magnet is practicable.

§ 9. Median Plane Measurement

In recent years, this measurement has become increasingly important in the design of charged particle accelerators which require the use of a magnetic guide field to keep the particles on a stable orbit between the walls of a vacuum chamber. The magnetic median plane can be defined as that plane in which particles will traverse an orbit without encountering any force which tends to displace them from it. It will be a plane where the component of the magnetic field normal to it is a maximum while the component in the plane is zero. Reference to figure 14 will show how a particle moving in this plane is subject only to a force in the plane, while a particle outside the plane will be subjected to a force tending to return it to or displace it from the plane, depending upon whether the field is focusing or defocusing. One method of median plane measurement has already

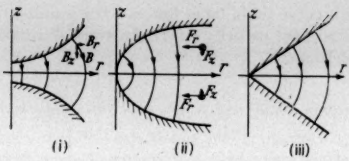


Figure 14. Showing three possible results of measuring vertical variation in the magnetic field using the nuclear resonance technique. The pole surfaces are shown shaded in each figure.

(i) $\frac{dB}{dr} < 0$; $\frac{d^2B}{dr^2} < 0$; maximum B on median plane;

(ii) $\frac{dB}{dr} < 0$; $\frac{d^2B}{dr^2} > 0$; minimum B on median plane;

(iii) $\frac{dB}{dr} < 0$; $\frac{d^2B}{dr^2} = 0$; equal values of B above, below and on the median plane.

been described in § 8.3. This is the 'floating-wire' method which depends on the forces acting on a current-carrying wire in a magnetic field.

The fact that there is a variation in the field component in or normal to the plane makes it possible to find the median plane in several ways. The first method involves the accurate determination of the variation of the field component normal to the magnetic median plane (this will be approximately the geometric median plane of the magnet pole faces generally). This may be done with any of the standard methods; for example, a coil with its plane parallel to the median plane may be moved on a line normal to the median plane and the answer obtained on a Grassot fluxmeter. In this case one looks for the position at which the maximum normal field component is obtained.

In a similar way, it is possible to estimate the position of the magnetic median plane using the technique of nuclear resonance, but one important feature should be remembered. The resonance frequency measured will be for the magnetic field in the direction of the lines of force and not the component normal to the magnetic median plane. Measuring the field in this way about the median plane can give a maximum, a minimum or no difference from other points. It is easily demonstrated that the result depends on d^2B/dr^2 if we consider an axially symmetric system, and this is shown in figure 14. In magnets for particle accelerators, the value of dB/dr is invariably negative so that there will be a minimum or maximum reading on, or equal readings above and below, the median plane depending on whether d^2B/dr^2 is greater than, less than or equal to zero. In all cases there does exist a median plane as far as focusing forces are concerned.

The third method involves the measurement of the component parallel to the median plane, which will go through zero at the median plane. In a magnet where the lines of force are vertical it is possible to use the force of gravity to define a direction. In a static field, the coil is suspended freely so that its axis is parallel to the median plane. In an axially symmetric system, the coil is set with its axis parallel to the radius vector and is then rotated through 180° and the reading on the Grassot fluxmeter taken. This is done for various vertical positions and the results plotted. The coil position at which the fluxmeter reading

is zero gives the position of the median plane. Where the radius of curvature of the magnetic lines of force is large, it is necessary to use this method since the vertical component changes very slowly in moving away from the median plane. This method also has the advantage that a zero is being sought.

For the case of dynamic fields, where the magnetic lines of force are vertical usually, the suspended coil can be readily used although in this instance it is necessary to make the measurement with the coil axis along the radius vector at different vertical positions and then to repeat the measurements with the coil rotated through 180°. To make the individual measurements, the voltage output from the coil, which is proportional to the time rate of change of the flux in the direction of the coil axis, may be integrated and the output of the integrator photographed from an oscilloscope. Because of the integration, it is necessary to allow for the initial condition of the field which may be measured by one of the techniques for static measurement. Plotting the two sets of results will give two curves which are in general displaced from each other since it is not possible to line the coil perfectly in a vertical position. However, the mean of these two curves will provide the necessary position where zero radial component exists. The technique is described by Richardson and Sewell (1951) with a description of the apparatus which they used and the accuracy achieved is a few tenths of a millimetre. A similar measurement is described by Green *et al.* (1953) where, instead of suspending the coil, it is mounted on a slab of machined marble which has been levelled in the magnet gap to a few seconds of arc. The remanent component of field in the radial direction is measured by rotating the coil on the level surface and taking the output to a fluxmeter. The dynamic measurements are made as above and the authors found that they could determine the position of the median plane to about $\pm 0{\cdot}5$ mm.

9.1. *Summary*

The technique of measuring the median plane of a magnet has become quite important and several methods are described which enable its position to be determined in both static and dynamic fields. The accuracy is of the order of a few tenths of a millimetre in general and for most purposes this is adequate. Apart from the 'floating-wire' technique mentioned in § 8, the other methods rely on a determination of the variation of either the normal component or the radial component parallel to the median plane.

§ 10. Miscellaneous Methods

10.1. *The Zeeman Effect*

This is an effect which is well known and treated in so many textbooks on atomic spectra (for example, White 1934) that it is better for the reader to investigate the literature. It is widely employed in spectral analyses and the magnetic field used may be measured by observing the effect in known lines of the spectra. For example, McNally and Harrison (1945) working on the Zeeman effect in the yttrium spectrum used lines in the spectra of silver, copper and calcium to estimate the field strength of their magnet, which in their case was as high as 84 000 gauss. Generally speaking the method is only applicable where considerable spectral resolving power is available.

An extension of the work on atomic spectra is to be found in the atomic beam magnetic resonance studies which were mentioned in § 2. This technique has been applied to some problems of a fundamental nature, such as the measurement of the hyperfine structure of the ground states of hydrogen and deuterium (Nafe and Nelson 1948, Prodell and Kusch 1952). To determine the intensity of the homogeneous field in which the transitions between magnetic levels occur, the frequency separation of the Zeeman components was measured. The fields used were from one gauss to several hundred gauss. This is a highly specialized field and the reader is advised to consult the paper of Nafe and Nelson for a comprehensive discussion of the subject. Two further articles of interest in this connection are those of Kellogg and Millman (1946) and Kusch (1951).

10.2. *The Magneto-Optical Rotation of the Plane of Polarization*

When polarized light is transmitted through a transparent medium which is subjected to a magnetic field, there is a rotation of the plane of polarization proportional to the length of the medium traversed and to the strength of the magnetic field. The constant of proportionality varies considerably for different substances and is usually known as Verdet's constant.

Using carbon bisulphide, Raoult (1949) has measured pulsed magnetic fields up to 50 000 gauss. The angle of rotation with a cell length of 8 mm was about 3°. A combination of a photocell and amplifier was used to measure the change in light intensity after the light had passed through the cell and an analysing Nicol prism. The accuracy achieved was to about 5%. Substances with somewhat greater Verdet constant exist and it would be possible to apply the technique to lower fields while maintaining a sufficiently large angle of rotation for reasonable measurement.

10.3. *Transmission of Light through a Colloidal Suspension*

When a liquid containing a suspension of small particles is subjected to a magnetic field, there is a change in the optical transmission of the medium. The effect was first observed in a qualitative way by Groves in 1845, and many papers on the subject have appeared since. With a ferromagnetic suspension, the theory remains incomplete but Stott (1949) has found it possible to develop a satisfactory theory for the effect using graphite, which is known to be diamagnetic and isotropic. Essentially the magnetic field exerts a torque on the small plates of graphite tending to align them perfectly. However, the particles of graphite form a true colloidal suspension and exhibit Brownian motion which tends to establish random orientation. The particles of graphite act like little blinds so that, as an increase of field strength lines them more into the direction of field on average, light transmitted parallel to the field will increase, while light at right angles to the field will decrease. The relation between the field strength and the amount of light transmitted is not a simple one and the system requires calibration if it is to be used as a field measuring technique. It would be useful in the range of fields up to about 2000 gauss. Papers by Stott (1949), von Engel (1950) and Stott and von Engel (1948) should be consulted for further information and references.

10.4. *The Magnetron Principle*

The magnetron in its simplest form consists of a cylindrical anode and a coaxial filament cathode in an evacuated envelope with a magnetic field applied parallel to the axis of the system. With a fixed voltage between the anode and cathode, the electrons move radially outwards if there is no magnetic field but in the presence of the magnetic field the paths become curved. At sufficiently large magnetic fields, the path becomes so curved that the electrons do not reach the anode. At this field value, the anode current should fall suddenly to zero. In practice, the cut-off is rather more gradual than this for several reasons and the current never quite reaches zero. Hull (1921, 1923) originally showed this condition could be predicted by the relation which is approximately

$$B_c = \frac{6{\cdot}74\, V^{1/2}}{d_1(1-d_2/d_1)}$$

where V is in volts, d_1 is the diameter of the anode in cm, and d_2 is the diameter of the cathode.

Gundlach (1938) and Weinzierl (1950) have described the use of such a tube for measuring magnetic fields. Using the cut-off point with constant anode current, a relation between B_c and V is obtained and this enables fields to the measured to 2%. A very small unit can be constructed and fields from about 100 to 10^4 gauss measured with ease. For the best results, care must be taken to line the cathode in the direction of the field. Where it is desired to measure variations around a given field value, the voltage may be held constant and, knowing the relation between the field and the anode current, field difference may be measured to 2%.

10.5. *The Mercury Jet Magnetometer*

Kolin (1945) has devised a measuring instrument which gives almost point measurement and has a well-defined direction of measurement associated with it. Mercury is pumped along a narrow channel of electrically insulating material into which are set two diametrically opposite electrodes, just projecting into the mercury in the channel. The motion of the mercury causes a voltage to be induced between the electrodes when there is a magnetic field with a component perpendicular to the axis of the channel. The voltage induced is given by

$$V = Bd\bar{v} \times 10^{-8} \text{ volts}$$

where B is the component of magnetic field perpendicular to the axis of the channel, in gauss d is the electrode separation in cm, and $\bar{v}$ the average velocity of the mercury in cm $\sec^{-1}$. Kolin states that an absolute determination of B is possible to 2%. Relative measurements may be made by the use of several probes while the relative values of the three rectangular components of B may be found by using a channel which traverses a path parallel to the three axes in turn, a pair of electrodes being set in each section of the path.

10.6. *Cathode-Ray Tube Devices*

The fact that electron paths are bent in a magnetic field has suggested a number of devices for various aspects of field measurement. Rose (1938) constructed a tube in which an accelerated beam of electrons traversed an electrical field free region in a gas pressure of several microns of argon. The gas served two purposes. It kept the beam of electrons well focused and it made the path

visible. By correctly orienting the tube so that the magnetic field was perpendicular to the direction in which the electron beam was ejected, the electrons traversed a circular path which returned to the axis of the electron gun again. The direction of the field was then the normal to the circular path while the magnitude was given by the same relation as for the magnetron cut-off, with the exception of the term involving the diameter of the cathode. Rose used this device over a range of fields from 1·5 to 300 gauss. In the range from 15 to 150 gauss, the accuracy of measurement lay in the region of a few per cent.

Kerst *et al.* (1950) describe a device which they term a ' zero field detector '. It consists of an electron gun giving a beam of 100 volt electrons which pass through a three-slit collimator into a Faraday cup. The beam could only enter the cup when the magnetic field was within plus or minus a quarter gauss of zero field. This was used to find the instant at which the field went through zero at different points around the gap of a betatron magnet.

10.7. *The Electrolytic Tank*

It is not intended to describe this technique here as it is adequately reviewed by Liebmann (1950) with which paper there is a very comprehensive list of references. The thesis of Bracher (1950), some of which is to be published shortly, contains a considerable amount of information on precision measurements using this technique.

§ 11. Conclusion

While an attempt has been made to cover the diverse range of techniques used in the measurement of fields, the main emphasis has been placed on those which have an important place in the everyday run of measurement. Where it has not been possible to give a full description, the reader should find it a comparatively easy task to extract that description from the references given.

Acknowledgments

The author wishes gratefully to acknowledge his indebtedness to the many people who have co-operated in producing this report by forwarding copies of published work, unpublished internal reports and notes. They are too numerous to mention in detail but their efforts are indicated by the large number of references collected at the end of this report. He is also indebted to Professor W. E. Burcham and Mr. A. Dee for their helpful criticism of the manuscript, and to the staff of the Harding Library, University of Birmingham, for their co-operation in searching through abstracts.

References

1. *Introduction*

Cotton, A., and Dupouy, G., 1932, *Congrès International d'Electricité, III, Sec.* 2*E*, *Report* 12.

McKeehan, L. W., 1929, *J. Opt. Soc. Amer.*, **19**, 213.

2. *Magnetic Resonance*

Beljers, H. G., 1953, *Philips Tech. Rev.*, **15**, 55.

Bleaney, B., and Stevens, K. W. H., 1953, *Rep. Progr. Phys.*, **16**, 108 (London: Physical Society).

BLOCH, F., 1946, *Phys. Rev.*, **70,** 460.
BLOCH, F., HANSEN, W. W., and PACKARD, M., 1946, *Phys. Rev.*, **69,** 127; *Ibid.*, **70,** 474.
BLOEMBERGEN, N., 1948, *Nuclear Magnetic Relaxation* (The Hague: Martinus Nijhoff).
BLOEMBERGEN, N., PURCELL, E. M., and POUND, R. V., 1948, *Phys. Rev.*, **73,** 679.
BRIGGS, G. R., 1950, *Internal Report, University of Illinois.*
DICKINSON, W. C., 1951, *Phys. Rev.*, **81,** 717.
DUMOND, J. W. M., BOGART, L., KOHL, J. L. MULLER, D. E., and WILTS, J. R., 1952, *California Instit. Tech. Report.*
EXTERMANN, DENIS and BENE, 1953, *Lectures at Conference on the alternating-gradient proton synchrotron at Geneva, IV*, 133.
GABILLARD, R., 1952, *Rev. Sci., Paris, No.* 3319.
GINDSBERG, J., and BEERS, Y., 1953, *Rev. Sci. Instrum.*, 24, 632.
GOLDSCHMIDT, S., 1920 *Ber. dtsch. chem. Ges.*, **53, 44**; 1922, *Ibid.*, **55,** 616; 1924, *Ann. Liebigs*, **437,** 137; *Ibid.*, **437,** 194.
GORTER, C. J., and BROER, L. J. F., 1942, *Physica*, **9,** 591.
GRIVET, P., and REGENSTREIF, E., 1953, *Lectures at Conference on the alternating-gradient proton synchrotron at Geneva, IV*, 127.
GUTOWSKY, H. S., MEYER, L. H., and MCCLURE, R. E., 1953, *Rev. Sci. Instrum.*, **24,** 644.
HOLDEN, A. N., KITTEL, C., MERRITT, F. R., and YAGER, W. A., 1949, *Phys. Rev.*, **75,** 1614; 1950, *Ibid.*, **77,** 147.
HOLDEN, A. N., YAGER, W. A., and MERRITT, F. R., 1951, *J. Chem. Phys.*, **19,** 1319.
HOPKINS, N. J., 1949, *Rev. Sci. Instrum.*, **20,** 401.
JACOBSOHN, B. A., and WANGSNESS, R. K., 1948, *Phys. Rev.*, **73,** 942.
KELLOGG, J. B. M., and MILLMAN, S., 1946, *Rev. Mod. Phys.*, **18,** 323.
KNOEBEL, and HAHN, 1951, *Rev. Sci. Instrum.*, **22,** 904.
KOSIREV, B. M., and SALEKOV, S. G., 1947, *Doklady Akad. Nauk*, **58,** 1023.
LINDSTRÖM, G., 1951, *Ark. Fys., B4, no.* 1; 1951: *Manne Siegbahn*, 1886–1951 (Uppsala: Almqvist and Wiksells).
PACKARD, M. E., 1948, *Rev. Sci. Instrum.*, **19,** 435.
POUND, R. V., 1947, *Phys. Rev.*, **72,** 527.
POUND, R. V., and KNIGHT, W. D., 1950, *Rev. Sci. Instrum.*, **21,** 219.
PURCELL, E. M., 1951, *Physica*, **17,** 282.
PURCELL, E. M., TORREY, H. C. and POUND, R. V., 1946, *Phys. Rev.*, **69,** 37.
RABI, J. J., MILLMAN, S., KUSCH, P., and ZACHARIAS, J. R., 1939, *Phys. Rev.*, **55,** 526.
RABI, J. J., ZACHARIAS, J. R., MILLMAN, S., and KUSCH, P., 1938, *Phys. Rev.*, **53,** 318.
RICHTER, H. L., HUMPHREY, F. B., and YOST, D. M., 1954, *Rev. Sci. Instrum.*, **25,** 190.
ROBERTS, A., 1947, *Rev. Sci. Instrum.*, **18,** 845.
ROBERTS, F. F., and JONES, G. O., 1951, *Post Office Engineering Department Research Report No.* 13509.
ROLLIN, B. V., 1948, *Rep. Progr. Phys.*, **12,** 22 (London: Physical Society).
TAUB, H., and KUSCH, P., 1949, *Phys. Rev.*, **75,** 1481.
THOMAS, H. A., 1952, *Electronics*, **25,** 114.
THOMAS, H. A., DRISCOLL, R. L., and HIPPLE, J. A., 1950, *J. Res. Nat. Bur. Stand.*, **44,** 569.
THOMAS, H. A., and HUNTOON, R. D., 1949, *Rev. Sci. Instrum.*, **20,** 516.
TOWNES, C. H., and TURKEVITCH, J., 1950, *Phys. Rev.*, **77,** 148.
TUTTLE, W. N., 1940, *Proc. Inst. Radio Engrs, N.Y.*, **28,** 23.
WARING, C. E., SPENCER, R. H., and CUSTER, R. L., 1952, *Rev. Sci. Instrum.*, **23,** 9.
ZAVOISKY, E., 1945, *J. Phys. U.S.S.R.*, **9,** 211, 245, 447.
ZUMMERMAN, J. R., and WILLIAMS, D., 1948, *Phys. Rev.*, **73,** 94.

3. *Magneto-resistance of Bismuth*

ALLEN, M., 1932, *Phys. Rev.*, **42,** 848.
BUBLITZ, G., 1938, *Arch. tech. Messen*, **83,** T61–3.
CONN, G. K. T., and DONOVAN, B., 1951, *J. Sci. Instrum.*, **28,** 7.
DONOVAN, B., and CONN, G. K. T., 1949, *Phil. Mag.*, **40,** 283.
HEURLINGER, T., 1916, *Phys. Z.*, **17,** 221.
International Critical Tables, 1926, **6,** 422.

KAPITZA, P., 1928, *Proc. Roy. Soc.* A, **119,** 358.
KELLER, H. B., 1953, *University of California Report UCRL–2249.*
KÖNIG, P. P., 1908, *Ann. Phys., Lpz.*, **25,** 921.
MACALPINE, W. W., 1931, *Phys. Rev.*, **37,** 624.
PETRESCU and LUCA, 1948, *Bull. Inst. Polytech. Jassy*, **3,** 808.
SCHNEIDER, G. W., 1928, *Phys. Rev.*, **31,** 251.
SMITH, G. S., 1937, *Elect. Engng, N.Y.*, **56,** 441, 475, 1400; 1939, *Trans. Amer. Inst. Elect. Engrs*, **58,** 52.
STIERSTADT, O., 1933, *Z. Phys.* **85,** 310; *Phys. Rev.*, **43,** 577.

4. *Hall Effect*

AIGRAIN, P., 1954, *Ann. Radioelect.*, **9,** 44.
BUBLITZ, G., 1938, *Arch. tech. Messen*, **83,** T61–3.
CUNNELL, F. A., SAKER, E. W. and EDMOND, J. T., 1953, *Proc. Phys. Soc.* A, **66,** 1115.
DONOGHUE, J. J. and EATHERLEY, W. P., 1951, *Rev. Sci. Instrum.*, **22,** 513.
DUNLAP, W. C., Jr., 1947, *Phys. Rev.* **71,** 471.
MASON, W. P., HEWITT, W. H., and WICK, R. F., 1953, *J. Appl. Phys.*, **24,** 166.
PEARSON, G. L., 1948, *Rev. Sci. Instrum.*, **19,** 263.
TANENBAUM, M., and MAITA, J. P., 1953, *Phys. Rev.*, **91,** 1009.
WEISS, H., 1953, *Z. Naturf.*, **8a,** 463.
WELKER, H., 1952, *Z. Naturf.*, **11,** 744.

5. *Peaking Strips*

ADAMS, G. D., DRESSEL, R. W. and TOWSLEY, F. E., 1950, *Rev. Sci. Instrum.*, **21,** 69.
ADAMS, G. D., KERST, D. and SCAG, D. T., 1947, *Rev. Sci. Instrum.*, **18,** 799.
BOSLEY, W., CRAGGS, J. D., and MCEWAN, D. H., 1947, *Nature, Lond.*, **159,** 229.
CLARK, A. F., 1946, *Phys. Rev.*, **70,** 444.
GIORDANO, S., GREEN, G. K., and ROGERS, E. J., 1953, *Rev. Sci. Instrum.*, **24,** 848.
GREGG, E. C., 1947, *Rev. Sci. Instrum.*, **18,** 77.
KELLY, J. M., 1951, *Rev. Sci. Instrum.*, **22,** 256.
KÜHNE, R., 1952, *Arch. tech. Messen*, **199,** 175.

6. *Generator Principle*

CALDECOURT, V. J., 1950, *Rev. Sci. Instrum.*, **21,** 772.
COLE, R. H., 1938, *Rev. Sci. Instrum.*, **9,** 215.
CORK, J. M., SCHREFFLER, R. G., and SHULL, F. B., 1947, *Rev. Sci. Instrum.*, **18,** 315.
DAYHOFF, E. S., TRIEBWASSER, S., and LAMB, W. E., Jr., 1953, *Phys. Rev.*, **89,** 106.
GODDARD, L. S., and KLEMPERER, O., 1944, *Proc. Phys. Soc.*, **56,** 378.
GRIVET, P., and REGENSTREIF, E., 1953, *Lectures at Conference on the alternating-gradient proton synchrotron at Geneva, IV*, 127.
HEDGRAN, A., 1952, *Ark. Fys.*, B5, *no.* 1; also 1951: *Manne Siegbahn*, 1886–1951 (Uppsala: Almqvist and Wiksells).
JÜRGENS, B. F., 1953, *Philips Tech. Rev.*, **15,** 49.
KATZ, L., FORSYTH, P. A., CUDNEY, L. F., WILLIAMS, G. W., JOHNS, H. E., and HASLAM, R. N. H., 1950, *Canad. J. Res. A*, **28,** 67.
KLEMPERER, O., and MILLER, H., 1939, *J. Sci. Instrum.*, **16,** 121.
KOHAUT, A., 1937, *Z. tech. Phys.*, **18,** 198.
KONAKHOVITCH, LATISCHEV and TSIMBALIN, 1949, *Izv. Akad. Nauk, ser. Fiz.*, **13,** 456.
KUNTZIGER, J., and LOUON, P., 1943, *Rev. Univ. Min.*, **86,** 211.
LAMB, W. E., Jr., and RETHERFORD, R. C., 1951, *Phys. Rev.*, **81,** 222.
LANGER, L. M., and SCOTT, F. R., 1950, *Rev. Sci. Instrum.*, **21,** 522.
PARKINSON, W. C., GROVER, G. M., and CRANE, H. R., 1947, *Rev. Sci. Instrum.*, **18,** 734.
PETERS, W. A. E., 1950, *Electrotech., Z.*, **71,** 193.
PLOCH, W., 1949, *Z. angew. Phys.*, **1,** 525.
STEINGROVER, E., 1952, *Arch. Elektrotech.*, **40,** 275.

TRNKA, Z., 1950, *Elekt. Obozr.*, **39,** 127.
VAN DER WALT, N. T., 1953, *Rev. Sci. Instrum.*, **24,** 413.

7. *Fluxmeter Methods*

BRIGGS, J. H., and MITCHELL, W. H., 1949, *J. Sci. Instrum.*, **26,** 40.
BROWN, W., and SWEER, J. H., 1945, *Rev. Sci. Instrum.*, **16,** 276.
CHANCE, B., 1949, *Waveforms, M.I.T. Rad. Lab. Series, Vol.* 19 (New York: McGraw-Hill), p. 664.
CIOFFI, P. P., 1950, *Rev. Sci. Instrum.*, **21,** 624.
COOK, D. B., and MATHER, R. L., 1947, *Internal Report, Columbia University.*
DICKE, R. H., 1948, *Rev. Sci. Instrum.*, **19,** 533.
DUNN, H. K., 1939, *Rev. Sci. Instrum.*, **10,** 368.
EDGAR, R. F., 1937, *Trans. Amer. Inst. Elect. Engrs*, **56,** 805.
EKELOF, S., BENGTSON, L., KIHLBERG, G., and LEITHAMMEL, P., 1951, *Chalmers University of Technology Translation, No.* 120.
FINLAY, E. A., FOWLER, J. F., and SMEE, J. F., 1950, *J. Sci. Instrum.*, **27,** 264.
FULLER, E. W., and HIBBARD, L. U., 1954, *J. Sci. Instrum.*, **31,** 36.
GREEN, G. K., KASSNER, R. R., MOORE, W. H., and SMITH, L. W., 1953, *Rev. Sci. Instrum.*, **24,** 743.
GREGG, E. C., 1949, *Rev. Sci. Instrum.*, **20,** 841.
HARRIS, L. H., 1931, *P.O. Elect. Engrs., J.* **24,** 206.
HERZOG, R. F. K., and TISCHLER, O., 1953, *Rev. Sci. Instrum.*, **24,** 1000.
INOZEMTCEV, K. V., and LATISHEV, G. D., 1949, *Izv. Akad. Nauk, ser. Fiz.* 13, 453.
KAUFMANN, E. N., 1950, *Radio and Television News*, **43,** 3, 20.
LAURITSEN, C. C., and LAURITSEN, T., 1948, *Rev. Sci. Instrum.*, **19,** 916.
MIKHAIL, H., and YOUSEF, Y. L., 1945, *Proc. Phys. Soc.*, **57,** 238.
MILNE, E. A., 1951, *Thesis, California Institute of Technology.*
NETTLETON, H. R., and SUGDEN, G., 1939, *Proc. Roy. Soc.* A, **173,** 314.
SPILSBURY, R. S. J., and WEBB, C. E., 1945, *J. Sci. Instrum.*, **22,** 213.
VERSTER, N. F., 1949, *Appl. Sci. Res., Hague, B,* **1,** 363.
WILLIAMS, A. J., Jr., 1938, *Trans. Amer. Inst. Elect. Engrs*, **57,** 565.
WILLIAMSON, K. J., 1947, *J. Sci. Instrum.*, **24,** 242.

8. *Forces on Conductors*

BIREBENT, R., 1952, *C.R. Acad. Sci., Paris*, **234,** 1135.
BRIGGS, G. H., 1936, *Proc. Roy. Soc.* A, **157,** 183.
BRIGGS, G. H., and HARPER, A. F. A., 1936, *J. Sci. Instrum.*, **13,** 119.
CHANG, W. Y., and ROSENBLUM, S., 1945, *Rev. Sci. Instrum.*, **16,** 75.
COTTON, A., and DUPOUY, G., 1932, *Congrès International d'Electricité, III, Sec.* 2E, *Report* 12.
CRANBERG, L., 1951, *U.S. Atomic Energy Commission Report AECU*–1670.
VAN DORSTEN, A. C., and FRANKEN, A. J. J., 1953, *Philips Tech. Rev.*, **15,** 52.
KAUFMANN, E. N., 1938, *Rev. Sci. Instrum.*, **9,** 369.
LOEB, J., 1947, *L'Onde Elect.*, **27,** 27.
MATES, J., 1951, *Carnegie Institute of Technology Report No. NYO*–778.
VAN MENTS, M., and LE POOLE, J. B., 1947, *Appl. Sci. Res., Hague,* **B1,** 3.
ROGERS, F. T., Jr., 1936, *Phys. Rev.*, **50,** 515; 1937, *Phys. Rev.*, **51,** 588.
SCOTT, F. A., 1934, *Phys. Rev.*, **46,** 633.
SLÄTIS, H., 1953, *Ark. Fys., B6,* 415.
THOMAS, H. A., DRISCOLL, R. L., and HIPPLE, J. A., 1950, *J. Res. Nat. Bur. Stand.*, **44,** 569.
UNIVERSITY OF CHICAGO, 1949, *Chicago Progress Report No.* 3.

9. *Median Plane Measurement*

GREEN, G. K., KASSNER, R. R., MOORE, W. H., and SMITH, L. W., 1953, *Rev. Sci. Instrum.*, **24,** 743.
RICHARDSON, R. E., and SEWELL, D. C., 1951, *Rev. Sci. Instrum.*, **22,** 697.

10. *Miscellaneous Methods*

Bracher, D., 1950, *Thesis*, University of Birmingham.

von Engel, A., 1950, *Arch. tech. Messen*, **179**, V391–6, T132.

Gundlach, F. W., 1938, *Arch. tech. Messen*, **85**, V391–3, 87.

Hull, A. W., 1921, *Phys. Rev.*, **18**, 31; 1923, *Ibid.*, **22**, 279.

Kellogg, J. B. M., and Millman, S., 1946, *Rev. Mod. Phys.*, **18**, 323.

Kerst, D., Adams, G. D., Koch, H. W., and Robinson, C. S., 1950, *Rev. Sci. Instrum.*, **21**, 462.

Kolin, A., 1945, *Rev. Sci. Instrum.*, **16**, 209.

Kusch, P., 1951, *Physica, Amsterdam*, **17**, 339.

Liebmann, G., 1950, *Advanc. Electron.*, **2**, 101.

McNally, J. R., and Harrison, G. R., 1945, *J., Opt. Soc. Amer.*, **35**, 584.

Nafe, J. E., and Nelson, E. B., 1948, *Phys. Rev.*, **73**, 718.

Prodell, A. G., and Kusch, P., 1952, *Phys. Rev.*, **88**, 184.

Raoult, G., 1949, *Ann. Phys., Paris*, **4**, 369.

Rose, A., 1938, *Electronics*, **11**, July, 21.

Stott, F. D., 1948, *Proc. Phys. Soc.* B, **62**, 1948.

Stott, F. D., and von Engel, A., 1948, *Nature, Lond.*, **161**, 728.

Weinzierl, P. M., 1950, *Rev. Sci. Instrum.*, **21**, 492.

White, H. E., 1934, *Introduction to Atomic Spectra* (New York: McGraw-Hill).

THEORY OF RADIATION

By J. C. GUNN

Natural Philosophy Department, University of Glasgow

CONTENTS

		PAGE
§ 1.	Introduction	128
§ 2.	Lagrangian and Hamiltonian formalism	130
	2.1. Particle mechanics	130
	2.2. Field mechanics	130
§ 3.	Scalar meson field	132
	3.1. Field equations	132
	3.2. Covariant commutation relationships	133
	3.3. Annihilation and creation operators	135
	3.4. The invariant Δ-functions	136
	3.5. Charged fields	137
§ 4.	Electromagnetic field	138
§ 5.	Electron–positron field	140
	5.1. Field equations	140
	5.2. Projection operators and spurs	142
	5.3. Quantization of electron–positron field	142
	5.4. Annihilation and creation operators	144
	5.5. Charge conjugation	145
	5.6. Vacuum expectation values	146
§ 6.	Field interactions	146
	6.1. Electromagnetic interactions	146
	6.2. Other interactions	147
§ 7.	Solution of the equations of field theory	148
	7.1. Schrödinger representation	148
	7.2. Interaction representation	149
	7.3. Dyson's solution for the S-matrix	151
	7.4. Evaluation of S-matrix elements	152
	7.5. Evaluation of cross sections	156
§ 8.	Elementary applications	158
	8.1. Compton scattering	158
	8.2. Møller scattering	159
	8.3. Scattering of light by light	160
§ 9.	Renormalization	161
	9.1. General survey	161
	9.2. Primitive divergents	163
	9.3. Subtraction of infinities	166
§ 10.	Calculation of radiative effects	171
	10.1. Introduction	171
	10.2. Electron self-energy	171
	10.3. Anomalous electron magnetic moment	172
	10.4. Radiative corrections to electron scattering	174
	10.5. The Lamb shift	175
	10.6. Compton scattering	180
§ 11.	Concluding remarks	181
	References	182

Abstract. Recent developments in the theory of radiation have brought this subject to a very complete state. There is a nearly closed domain, in which the interaction between charged particles and the electromagnetic field can be considered as isolated from other phenomena such as mesonic interactions. Within this domain it is now possible to calculate any process with a precision only limited by the labour of the actual calculations. This progress requires for its achievement that quantum field theory should be developed in a relativistically covariant form. The early sections of this review are devoted to a systematic description of the methods of quantization as applied to various matter fields—the scalar meson field considered first for its simplicity, the electromagnetic field and the electron–positron field. Coupled field equations are then discussed, and their solution in the interaction representation is described. The Feynman–Dyson procedure is established for the evaluation of S-matrix elements and applied to typical cases. A general description is given of different methods of renormalization of the S-matrix, and it is shown how this is equivalent to an elimination of unobservable parameters, and a description of phenomena in terms of the (formally infinite) electron mass and charge. A series of individual problems, exhibiting radiative corrections, including the Lamb shift and the anomalous magnetic moment of the electron are considered in more or less detail. Some brief concluding remarks are appended on the development of approximations likely to be useful in meson theory.

§ 1. Introduction

The development of fundamental particle physics has in many ways been the central theme of physics during the last thirty years. The conception that matter (or energy) always has its existence in one of the forms described by a few fundamental field equations indeed seems a very simple one. This was certainly true so long as the only elementary particles one had to deal with were the nucleons, the electrons and positrons, the electromagnetic field quanta and perhaps one or two varieties of meson. Of recent years with the experimental discovery of a host of new mesons and hyperons one is perhaps more doubtful how far it will prove correct to describe physical reality in terms of the superposition of a great set of fundamental particle fields coupled together in a set of linear equations. One can, however, be in no doubt as to the validity of the application of the methods of quantum field mechanics to the theory of radiation—the interaction of charged particles with the electromagnetic field. The theory of radiation can never itself be a complete theory. Even in processes involving only photons and electrons there is some influence due to coupling, for example, with the meson field. However, in large measure such extraneous interactions can be disregarded, and we deal with a very nearly closed system. It is well known that the theory of radiation cannot be developed without involving some infinities, and it is perhaps desirable to leave open the possibility that these are in some way removed when all natural fields are considered together. At the same time, relations might be forthcoming which would limit the apparently large number of constants of nature needed to describe all the observed 'fundamental' particles and the couplings between their fields.

The programme of this report then is to consider the fundamental interactions between charged particles and the electromagnetic field.† It was in connection with such interactions that the idea of field quantization was first introduced—

† For other fields reference may be made to an article "Survey of Field Theory" by R. E. Peierls *et al.* in this volume.

for the radiation field by Dirac (1927), for the electron field by Jordan and Wigner (1928), for electrodynamics by Heisenberg and Pauli (1929). The ideas were very rapidly developed and applied to a wide variety of problems with great success. The formal state of the theory in 1940 is well described in a review article by Pauli (1941), and the application to individual problems is fully given in the earlier editions of Heitler's well-known book on the *Quantum Theory of Radiation.* There was a second wave of advance in the subject in the years 1946–1949. This was mainly due to the work of Tomonaga, Schwinger, Feynman and Dyson, more detailed reference to which is given in later sections. These authors formulated the Schrödinger equation, describing the development of the wave function for a system of electrons and photons in interaction, in a relativistically invariant way, making use of the so-called interaction representation. The result has been, on the one hand, to lead to a simplification in the application of the theory to higher order processes, particularly using the methods of Feynman. On the other hand it has led to a method for the subtraction of the divergences from the calculation of these higher order effects in a covariant way. These subtraction processes were shown to be equivalent with a renormalization of the electron mass and charge, Dyson finally demonstrating that the procedure of subtraction of all the infinities was identical with a redefinition of the electron mass and charge.

The renormalized electrodynamics is capable, in principle, of yielding a result for any process involving the interaction of electrons and photons to an arbitrary degree of precision. Some of its applications are discussed in § 10 below. The ability to obtain results with such precision rests on the smallness of the fine structure constant $e^2/4\pi\hbar c = 1/137$. The expansion of the transition matrix element for any process, according to perturbation theory, in powers of e^2 will yield a series in which for the first few terms at least each term is very small compared with the preceding one. Physical arguments due to Dyson (1952) suggest that the series is in fact divergent, though it is presumably asymptotic in character. The exact relationship of the series to the true result depends on the unsolved question whether the equations of quantum electrodynamics, as at present formulated, in fact possess finite solutions.

The methods of radiation theory are, in principle, applicable also to other interacting fields such as mesons and nucleons. However, in this case the coupling constant is no longer small, and a perturbation expansion is quite inadequate. Various attempts are currently being made to find approximate solutions of the field equations not dependent on the smallness of the coupling constant. However, up to the present none of these has as yet proved completely successful. It has accordingly been thought best in this review largely to avoid discussion of such methods, though a few developments beyond the perturbation theory formalism, sufficient for the discussion of electrodynamics, are described in § 11.

Even if this restriction is accepted, the subject of radiation theory remains a large one, and it will be impossible to aim at a comprehensive description within this article. Such a description is less necessary in view of the recent edition by Heitler (1954) of his book on radiation theory. Rather the aim will be, within a relatively short compass, to give a systematic account of the quantization of fields and of the derivation of the renormalized S-matrix together with some discussion of the predicted effects due to radiative corrections.

§ 2. Lagrangian and Hamiltonian Formalism

2.1. *Particle Mechanics*

It will be useful first briefly to recall the Lagrangian formalism for classical particle mechanics and its relation to quantum mechanics. If we form the action integral

$$I = \int_{t_0}^{t_1} L(q, \dot{q})\, dt \qquad \ldots\ldots(2.1)$$

where L is the Lagrangian and q is used to denote the coordinates of the system, then Hamilton's principle states that I is stationary along the dynamical trajectory for variations of the q such that $\delta q(t_0) = \delta q(t_1) = 0$. In fact forming δI we have immediately

$$\delta I = \int_{t_0}^{t_1} \left(\frac{\partial L}{\partial q} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}}\right) \delta_0 q\, dt + \left[\frac{\partial L}{\partial \dot{q}}\, \delta_0 q\right]_{t_0}^{t_1} \qquad \ldots\ldots(2.2)$$

which leads to the Lagrange equations of motion ($\delta_0 q$ is used to denote the variation of a typical coordinate at a given time).

We consider now a more general variation (but from a dynamical trajectory) in which the initial and final times are also altered, say by an amount δt, then

$$\delta I = \left[L\delta t + \frac{\partial L}{\partial \dot{q}}\, \delta_0 q\right]_{t_0}^{t_1} = \left[\left(L - \dot{q}\frac{\partial L}{\partial \dot{q}}\right)\delta t + \frac{\partial L}{\partial \dot{q}}\, \delta q\right]_{t_0}^{t_1} \qquad \ldots\ldots(2.3)$$

where now the total variation of q at the end points is denoted by $\delta q = \delta_0 q + \dot{q}\, \delta t$. The Hamiltonian function H and the momentum p conjugate to q are then defined as the coefficients of $-\delta t$ and δq in I, viz.:

$$H = \dot{q}\frac{\partial L}{\partial \dot{q}} - L, \qquad p = \frac{\partial L}{\partial \dot{q}}. \qquad \ldots\ldots(2.4)$$

The commutation relations and the Heisenberg equations of motion of quantum mechanics are defined in terms of these quantities according to

$$[q_i, p_j] = i\,\delta_{ij} \quad ; \quad [q_i, q_j] = [p_i, p_j] = 0 \qquad \ldots\ldots(2.5)$$

$$i\frac{d\alpha}{dt} = [\alpha, H],$$

where α is any dynamical variable not involving the time explicitly. Natural units will be used throughout in which $\hbar = c = 1$.

2.2. Field Mechanics

The field requires for its specification a set of field functions $\psi_\sigma(\mathbf{x}, t) = \psi_\sigma(x)$. The finite number of coordinates q_i of particle mechanics is replaced by the continuous infinity $\psi_\sigma(\mathbf{x})$, the behaviour of which as time varies is to be found. As always here we denote by $x = (x_1, x_2, x_3, x_4)$ a general world point with $x_4 = ix_0 = it$; $\mathbf{x} = (x_1, x_2, x_3)$ denotes position in three-dimensional space. If a Lagrangian density exists defined by

$$\mathcal{L}(x) = \mathcal{L}\left(\psi_\sigma(x), \frac{\partial\psi_\sigma(x)}{\partial x_\mu}\right)$$

then at any time the Lagrangian function is given by

$$L(t) = \int \mathcal{L}(x)\, d\mathbf{x}$$

and the corresponding classical equations of motion can be derived from variation of the action integral

$$I = \int_{t_0}^{t_1} L(t)\, dt = \int_{t_0}^{t_1} \mathcal{L}(x)\, dx$$

where $dx = dx_0\, dx_1\, dx_2\, dx_3$. If $\mathcal{L}$ is a scalar the equations will be of relativistically invariant form.

By analogy with equations (2.5) the corresponding quantum equations and commutation relations may also be derived, when the variation of I is chosen to involve also a variation of the boundaries of the space–time region over which $\mathcal{L}$ is integrated. To accord strictly with equation (2.3) this region of space–time should be a cylindrical one with generators parallel to the x_4, imaginary time, axis. This restriction is clearly against the requirements of relativistic invariance and can be abandoned. In general, if R is any space–time region and

$$I = \int_R \mathcal{L}(x)\, dx$$

then, if a variation $\delta_0\psi_\sigma(x)$ is made of the field functions together with a rigid body displacement δx_μ of the region R, we find

$$\delta I = \int_R \left\{ \frac{\partial \mathcal{L}}{\partial \psi_\sigma} - \frac{\partial}{\partial x_\nu} \frac{\partial \mathcal{L}}{\partial \psi_{\sigma,\nu}} \right\} \delta_0\psi_\sigma\, dx$$

$$+ \delta x_\mu \int_R \frac{\partial}{\partial x_\nu} \left\{ \mathcal{L}\delta_{\mu\nu} - \psi_{\sigma,\mu} \frac{\partial \mathcal{L}}{\partial \psi_{\sigma,\nu}} \right\} dx + \int_R \frac{\partial}{\partial x_\nu} \left\{ \frac{\partial \mathcal{L}}{\partial \psi_{\sigma,\nu}} \delta\psi_\sigma \right\} \quad \ldots\ldots (2.6)$$

(here $\psi_{\sigma,\nu} = \partial\psi_\sigma/\partial x_\nu$).

The vanishing of the first integral yields the field equations corresponding to the given Lagrangian

$$\frac{\partial \mathcal{L}}{\partial \psi_\sigma} = \frac{\partial}{\partial x_\nu} \frac{\partial \mathcal{L}}{\partial \psi_{\sigma,\nu}}.$$

The remaining integrals are of a four-dimensional divergence, and can be converted into surface terms. In fact if a four-dimensional surface element is defined by

$$d\sigma_\lambda = \frac{1}{i} (dx_2\, dx_3\, dx_4, \quad \ldots\ldots \quad, dx_1\, dx_2\, dx_3) \quad \ldots\ldots (2.8)$$

it follows that $\int (\partial A_\lambda/\partial x_\lambda) dx = \int A_\lambda d\sigma_\lambda$. In the present case, if the variation is made from a configuration satisfying the field equations

$$\delta I = \int d\sigma_\nu \left[\delta x_\mu \left(\mathcal{L}\delta_{\mu\nu} - \psi_{\sigma,\mu} \frac{\partial \mathcal{L}}{\partial \psi_{\sigma,\nu}} \right) + \delta\psi_\sigma \frac{\partial \mathcal{L}}{\partial \psi_{\sigma,\nu}} \right] \quad \ldots\ldots (2.9)$$

where $\delta\psi_\sigma = \delta_0\psi_\sigma + \psi_{\sigma,\mu}\,\delta x_\mu$ represents the total variation of ψ_σ between corresponding points of two boundaries. We limit consideration now to cylindrical regions of space–time but the boundaries need no longer be planes t = constant, but rather planes with normals in an arbitrary time-like direction n_μ ($n_\mu{}^2 = -1$) so that $d\sigma_\mu = n_\mu\, d\sigma$. For a plane t = constant, $n_\mu = (0, 0, 0, -i)$ and $d\sigma = dx_1\, dx_2\, dx_3$. We then find

$$\delta I = F(\sigma_1) - F(\sigma_0)$$

where $$F(\sigma_1) = \int_{\sigma_1}\left[\left(\frac{\partial \mathcal{L}}{\partial\psi_{\sigma,\nu}}n_\nu\right)\delta\psi_\sigma + \left(\mathcal{L}\delta_{\mu\nu} - \psi_{\sigma,\mu}\,\frac{\partial \mathcal{L}}{\delta\psi_{\sigma,\nu}}\right)n_\nu\,\delta x_\mu\right]d\sigma. \quad ..(2.10)$$

We now define $\pi_\sigma = n_\nu\dfrac{\partial \mathcal{L}}{\delta\psi_{\sigma,\nu}}$ and $P_\mu(\sigma) = \displaystyle\int_\sigma\left(\mathcal{L}\delta_{\mu\nu} - \psi_{\sigma,\mu}\,\frac{\partial \mathcal{L}}{\partial\psi_{\sigma,\nu}}\right)n_\nu\, d\sigma.$

By analogy with the rules for particle mechanics the commutation relations and Heisenberg equations of motion are, respectively,

$$\left[\psi_\sigma(x)\,,\int_{\sigma\ \text{thro' }x}\pi_\lambda(x')\,\delta\psi_\lambda(x')\,d\sigma'\right] = i\,\delta\psi_\sigma(x)$$

and
$$i\,da/dx_\mu = [P_\mu(\sigma),\, a]. \qquad .....(2.11)$$

The usual canonical formalism is found when σ is taken as a surface t = constant, which yields $\pi_\sigma = \partial\mathcal{L}/\partial\dot{\psi}_\sigma$ and

$$[\psi_\alpha(\mathbf{x}, t)\,,\pi_\beta(\mathbf{x}', t)] = i\,\delta_{\alpha\beta}\,\delta(\mathbf{x}-\mathbf{x}'),$$

$$[\psi_\alpha(\mathbf{x}, t)\,,\psi_\beta(\mathbf{x}', t)] = [\pi_\alpha(\mathbf{x}, t)\,,\pi_\beta(\mathbf{x}', t)] = 0. \quad .....(2.12)$$

In this case $P_\mu(\sigma)$ is the energy momentum four-vector of the system $P_\mu = (\mathbf{G}, iH)$.

§ 3. Scalar Meson Field

3.1. *Field Equations*

As the simplest example of the above formalism we treat briefly the case of a neutral scalar meson field ϕ, say. The Lagrangian density for this case may be written

$$\mathcal{L} = -\tfrac{1}{2}\left[\left(\frac{\partial\phi}{\partial x_\nu}\right)^2 + \mu^2\phi^2\right] \qquad .....(3.1)$$

(the most general relativistically invariant Lagrangian that can be constructed for a scalar field to yield a linear second-order equation). The field equation is immediately derived to be of the form

$$(\Box - \mu^2)\phi = 0 \qquad ....(3.2)$$

where as usual we denote $\Box = \partial_\nu{}^2 = \nabla^2 - \dfrac{\partial^2}{\partial t^2}$.

In the non-covariant canonical formalism $\pi(x) = \dot{\phi}(x)$ and the Hamiltonian and momentum densities are of the form

$$\mathcal{H} = \tfrac{1}{2}\{\pi^2 + (\nabla\phi)^2 + \mu^2\phi^2\}, \quad \mathcal{G}_k = -\pi\,\partial\phi/\partial x_k. \qquad .....(3.3)$$

For future purposes it will be necessary to know also the covariant commutation relationships, typically $[\phi(x)\ ,\ \phi(x')]$ where $x\ ,\ x'$ are connected by a time-like interval. Clearly the value of such a commutator is fully determined by the equation of motion together with the canonical commutation relations. Recently, however, an elegant method has been found by Peierls (1952) for deriving such covariant commutation relationships out of a knowledge of the Heisenberg equations of motion together with the field equations (these two of course imply a knowledge of the canonical relations). The method of Peierls has the advantage that it can be directly applied to Fermi–Dirac fields for which the canonical relations are inapplicable.

3.2. *Covariant Commutation Relationships*

Suppose it is desired to find the commutation relationship between two functions A, B of the field variables, $[A(x_1),\ B(x_2)]$. Peierls notes certain restrictions on these functions for his method to apply, which will certainly be satisfied in the particular cases considered later. Consider first a modified Lagrangian density

$$\mathcal{L}'(x) = \mathcal{L}(x) + \lambda A(x_1)\ \delta(x - x_1) \qquad \ldots\ldots(3.4)$$

where $\delta(x-x_1) = \delta(\mathbf{x}-\mathbf{x}_1)\ \delta(t-t_1)$. This implies a Lagrangian

$$L'(t) = \mathrm{L}(t) + \lambda A(x_1)\ \delta(t-t_1) \qquad \ldots\ldots(3.5)$$

which will yield modified field equations. In particular the new solution for the variable $B(x)$ will be written in the form

$$B'(x) = B(x) + \lambda\ \delta_1 B(x). \qquad \ldots\ldots(3.6)$$

$B'(x)$ is completely defined by the equation of motion, together with the condition $\delta_1 B \to 0$ as $t \to -\infty$.

If A contains no time derivatives of field quantities the modified Hamiltonian of the system will be

$$H'(t) = H - \lambda\ A(x_1)\ \delta(t-t_1) = H - \delta H. \qquad \ldots.(3.7)$$

It can now be shown that $B'(t)$, which is a solution of the Heisenberg equation of motion with Hamiltonian H', and $B(t)$ are related by a transformation

$$B'(t) = S(t)\ B(t)\ S^{-1}(t)\ ,\ \text{with}\quad i\,\partial S/\partial t = \delta H\, S. \qquad \ldots\ldots(3.8)$$

To the first order in λ the solution for S can be written

$$S(t) = 1 - i\lambda\ \theta(t-t_1)\ A(x_1), \quad \theta(t) = 0, \quad t < 0 \qquad \ldots\ldots(3.9)$$
$$= 1, \quad t > 0.$$

Using this expression in equation (3·8) and comparing with equation (3·6) we find

$$\delta_1 B(x_2) = \frac{1}{i}\ [A(x_1),\ B(x_2)]\ \theta(x_2 - x_1).$$

Similarly $\quad \delta_2 A(x_1) = -\frac{1}{i}\ [A(x_1),\ B(x_2)]\ \theta(x_1 - x_2).$

Combining these leads to a symmetrical form of the desired result

$$[A(x_1),\ B(x_2)] = i\{\ \delta_1 B(x_2) - \delta_2 A(x_1)\ \}. \qquad \ldots\ldots(3.10)$$

This relationship can immediately be applied to the calculation of the commutator for the scalar meson field $[\phi(x_1)\ ,\ \phi(x_2)]$. First, to evaluate $\delta_1\phi(x_2)$ we must take

$$\mathcal{L}'(x) = \mathcal{L}(x) + \lambda\ \delta(x - x_1)\ \phi(x).$$

The equation determining $\delta_1\ \phi$ is then

$$(\Box - \mu^2)\ \delta_1\ \phi(x) = -\delta(x - x_1)\ ,\ \delta_1\phi = 0 \text{ for } t < t_1 \quad \ldots\ldots(3.11)$$

and the solution with the appropriate boundary condition is

$$\delta_1\ \phi(x) = \Delta_R\ (x - x_1)$$

where Δ_R is the retarded Green's function for the equation (3.2) i.e. it satisfies

$$(\Box - \mu^2)\ \Delta_R\ (x) = -\delta(x) \text{ with } \Delta_R(x) = 0,\ t < 0. \ldots\ldots(3.12)$$

Using the expansion $\delta(x) = \left(\frac{1}{2\pi}\right)^4 \int e^{ikx}\, d^4k,$

where kx denotes the scalar product $k_\mu x_\mu = \mathbf{k}\ .\ \mathbf{x} - k_0\ x_0$, $\Delta_R(x)$ can be expressed

$$\Delta_R(x) = \left(\frac{1}{2\pi}\right)^4 \int_{C_R} \frac{e^{ikx}}{k^2 + \mu^2} d^4k. \quad \ldots\ldots(3.13)$$

Where there is no risk of confusion d^4k will often be replaced by dk. The contour C_R has to be chosen to satisfy the boundary condition (3.12). The integrand of equation (3.13) has poles where $k_0 = \pm\sqrt{(\mathbf{k}^2 + \mu^2)} = \pm\kappa$ say. If the path of integration over k_0 is that denoted by C_R in figure 1, then it is immediately seen that for $t < 0$ the contour can be deformed into an infinite semi-circle in the upper half of the k_0-plane along which the integrand vanishes exponentially so that the integral becomes zero.

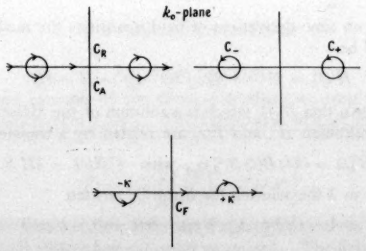


Figure 1. Contours for evaluation of Δ- and S-functions.

From equation (3.10) it now follows that

$$[\phi(x)\ ,\ \phi(y)] = i\ [\Delta_R(y - x) - \Delta_R(x - y)] \quad \ldots\ldots(3.14)$$

We may introduce the advanced Green's function $\Delta_A(x)$ defined by (3.12) together with $\Delta_A\ (x) = 0,\ t > 0$. $\Delta_A\ (x)$ has the form (3.13) where the contour C_A now runs below the singularities in the k_0-plane. It is immediately seen that $\Delta_A(x) = \Delta_R(-x)$. The Δ function is defined by

$$\Delta(x) = \Delta_A(x) - \Delta_R(x). \qquad \ldots\ldots(3.15)$$

This satisfies the field equation without sources, and can be written in the form

$$\Delta(x) = \left(\frac{1}{2\pi}\right)^4 \int_C \frac{e^{ikx}}{k^2 + \mu^2}\, dk. \qquad \ldots\ldots(3.16)$$

The contour C consists of two circles enclosing the poles, described in an anti-clockwise direction. The k_0-integration can immediately be carried out in this case yielding

$$\Delta(x) = \frac{-i}{(2\pi)^3}\int \frac{e^{ikx}}{2\kappa}\, [\delta(k_0 - \kappa) - \delta(k_0 + \kappa)]\, dk$$

$$= \frac{-i}{(2\pi)^3}\int e^{ikx}\ \delta(k^2 + \mu^2)\ \epsilon(k)\, dk \qquad \ldots\ldots(3.17)$$

where
$$\epsilon(k) = +1 \quad k_0 > 0$$
$$\quad\ -1 \quad k_0 < 0.$$

In terms of the Δ -function

$$[\phi(x)\,,\ \phi(y)] = i\,\Delta(x - y). \qquad \ldots\ldots(3.18)$$

3.3. *Annihilation and Creation Operators*

It remains to be seen how the field quantization of the last few paragraphs leads to the equivalence of the quantized field with an assembly of particles. If we expand the operator $\phi(x)$ in terms of the plane wave solutions of equation (3.2), in view of the reality of $\phi(x)$ we obtain an expression of the form

$$\phi(x) = \int \frac{dk}{(2\pi)^{3/2}} \frac{1}{(2\kappa)^{1/2}} \left[a_k\, e^{ikx} + a_k^*\, e^{-ikx}\right] \delta(k_0 - \kappa). \qquad \ldots\ldots(3.19)$$

With the help of the invariant commutation relations it is easy to show

$$[a_k\,,\ a_{k'}] = 0 = [a_k^*,\ a_{k'}^*]\ ,\ \ [a_k\,,\ a_{k'}^*] = \delta(\mathbf{k} - \mathbf{k}'). \qquad \ldots\ldots(3.20)$$

Sometimes in place of this continuum expansion it is convenient to choose a normalization volume V. The permitted $\mathbf{k}$ then become a discrete series, and the same commutation relations as above are obtained, with the replacement of the δ-function by a Kronecker δ-symbol if we make the substitution

$$\int \frac{d\mathbf{k}}{(2\pi)^{3/2}} \longrightarrow \frac{1}{V^{1/2}} \sum_{\mathbf{k}}.$$

The Hamiltonian H may be calculated from equation (3.3) and takes the form

$$H = \int d\mathbf{k}\ \ a_k^* a_k \kappa \qquad \ldots\ldots(3.21)$$

from which, or directly from the Heisenberg equation for ϕ, it may be deduced that

$$\kappa a_k = [a_k\,,\ \mathrm{H}]; \quad -\kappa a_k^* = [a_k^*\,,\ \mathrm{H}]. \qquad \ldots\ldots(3.22)$$

The vector character of the suffix k is henceforward ignored to simplify appearances. Consider now the operator

$$N = \int a_k^*\, a_k\, d\mathbf{k} = \int N_k\, d\mathbf{k}. \qquad \ldots\ldots(3.23)$$

It may easily be shown that this operator can only have integral eigenvalues, which must all be positive or zero. Suppose in particular that Ψ_n is an eigenstate belonging to the value n, so that $N\Psi_n = n\,\Psi_n$. Then from the commutation relations it follows directly that

$$N\,a_k\,\Psi_n = (n-1)\,a_k\,\Psi_n. \qquad \ldots\ldots(3.24)$$

The lack of negative eigenvalues shows that the relations of the sort (3.24) must terminate with the eigenvalue zero for N: $N\,\Psi_0 = 0$. Ψ_0 is then the vacuum state which can be defined by the relation $a_k\,\Psi_0 = 0$.

Equation (3.24) also leads to the interpretation of a_k as an operator for the annihilation of a particle, N_k as the total number of particles. From equation (3.21) and the corresponding expression $\mathbf{G} = \int N_k\,\mathbf{k}\,d\mathbf{k}$ for the field momentum, it appears that a_k annihilates a particle of momentum $\mathbf{k}$ and energy κ. a_k^* acts as the corresponding creation operator.

3.4. *The Invariant Δ–Functions*

There is an important series of Δ–functions related to the vacuum expectation values of various field operators which may conveniently be included at this stage. These are given for the case of the neutral scalar meson field but exactly analogous functions arise for the electromagnetic and electron–positron fields. The vacuum state is defined by the property $a_k\,\Psi_0 = 0$, so that also $\Psi_0\,a_k^* = 0$. It follows immediately that the only non-vanishing expectation value in the vacuum state for an operator bilinear in the a_k, a_k^* is

$$\langle a_k\,a_{k'}{}^*\rangle_0 = \delta(\mathbf{k}-\mathbf{k}'). \qquad \ldots\ldots(3.25)$$

Consider now $\langle\phi(x)\phi(y)\rangle_0$. Using the expression (3.19) for ϕ it will be seen that

$$\langle\phi(x)\,\phi(y)\rangle_0 = i\,\Delta^+(x-y) \qquad \ldots\ldots(3.26)$$

where $\Delta^+(x)$ is the positive frequency part of $\Delta(x)$, i.e. that part contributed by the contour C^+, figure 1. $\Delta^-(x)$ is similarly defined so that $\Delta(x) = \Delta^+(x) + \Delta^-(x)$. Then $\Delta^{(1)}(x)$ is defined by

$$-i\,\Delta^{(1)}(x) = \Delta^+(x) - \Delta^-(x). \qquad \ldots\ldots(3.27)$$

From the definition of Δ^+ and Δ^- we see that $\Delta^+(-x) = -\Delta^-(x)$ which leads to

$$\langle\{\,\phi(x)\,,\,\phi(y)\,\}\rangle_0 = \Delta^{(1)}(x-y). \qquad \ldots\ldots(3.28)$$

Here the curly bracket denotes the anti-commutator

$$\{\,\phi(x)\,,\,\phi(y)\,\} = \phi(x)\,\phi(y) + \phi(y)\,\phi(x).$$

The even Green's function $\overline{\Delta}(x)$ is defined as the mean of the advanced and retarded Δ–functions

$$\overline{\Delta}(x) = \tfrac{1}{2}\{\,\Delta_R(x) + \Delta_A(x)\,\} = -\tfrac{1}{2}\,\epsilon(x)\,\Delta(x). \qquad \ldots\ldots(3.29)$$

Finally the Feynman Δ–function is defined by

$$\Delta_F(x) = \Delta^{(1)}(x) - 2i\,\overline{\Delta}(x). \qquad \ldots\ldots(3.30)$$

This is also a Green's function having the property $(\Box - \mu^2)\,\Delta_F(x) = 2i\delta(x)$. Most important for our purposes is that it describes the vacuum expectation value of the time-ordered product $P(\phi(x), \phi(y))$. This notation due to Dyson

means that the product is to be so ordered that the operator on the right corresponds to the world point with the earliest time, i.e.

$$P(\phi(x), \phi(y)) = \begin{cases} \phi(x)\,\phi(y) & \text{if } y_0 < x_0 \\ \phi(y)\,\phi(x) & \text{if } x_0 < y_0. \end{cases} \qquad \ldots\ldots(3.31)$$

We can write

$$P(\phi(x), \phi(y) = \tfrac{1}{2}\{\phi(x), \phi(y)\} + \tfrac{1}{2}\,\epsilon(x-y)\,[\phi(x), \phi(y)]$$

so that

$$\langle P(\phi(x), \phi(y))\rangle_0 = \tfrac{1}{2}\Delta_F(x-y). \qquad \ldots\ldots(3.32)$$

Δ_F can again be expressed in the form of a contour integral

$$\Delta_F = -2i\left(\frac{1}{2\pi}\right)^4 \int_{C_F} \frac{e^{ikx}\,dk}{k^2 + \mu^2} \qquad \ldots\ldots(3.33)$$

where C_F is indented below the singularity at $k_0 = -\kappa$, and above that at $k_0 = \kappa$. This means that $\Delta_F(x)$ contains only positive frequencies for $t > 0$ and only negative frequencies for $t < 0$.

3.5. *Charged Fields*

Out of a single neutral field it is impossible to construct a current density vector j_μ satisfying the continuity equation $\partial j_\mu/\partial x_\mu = 0$. Charged fields become possible, however, when the field function is complex and the Lagrangian density $\mathcal{L}(x)$ is invariant to the transformation

$$\phi \to \phi\, e^{i\alpha} \quad , \quad \phi^* \to \phi^*\, e^{-i\alpha}. \qquad \ldots\ldots(3.34)$$

The complex field ϕ is equivalent to two real fields ϕ_1, ϕ_2 defined by $\phi = (\phi_1 + i\phi_2)/\sqrt{2}$, $\phi^* = (\phi_1 - i\phi_2)/\sqrt{2}$. In carrying out the variational procedures of § 2.2 it follows that ϕ, ϕ^* can be regarded as capable of independent variation. It may be noted that the gauge transformation (3.34) with respect to which $\mathcal{L}$ must be invariant is equivalent to a rotation through an angle α in the (12) plane when ϕ_1, ϕ_2 are regarded as components of a vector in an abstract (isotopic) space.

The current density vector has the form

$$j_\nu = -ie\left(\frac{\partial \mathcal{L}}{\partial \phi_\nu}\,\phi - \frac{\partial \mathcal{L}}{\partial \phi_\nu^*}\,\phi^*\right) \qquad \ldots\ldots(3.35)$$

where as usual $\phi_\nu = \partial\phi/\partial x_\nu$.

For the complex field the field equations have the form analogous to equation (2.7)

$$\frac{\partial \mathcal{L}}{\partial \phi} = \frac{\partial}{\partial x_\nu}\frac{\partial \mathcal{L}}{\partial \phi_\nu} \quad ; \quad \frac{\partial \mathcal{L}}{\partial \phi^*} = \frac{\partial}{\partial x_\nu}\frac{\partial \mathcal{L}}{\partial \phi_\nu^*}. \qquad \ldots\ldots(3.36)$$

Also in the canonical formalism the momentum π conjugate to ϕ (on a plane t = constant) is given by $\pi = \partial \mathcal{L}/\partial\dot{\phi}$, and the Hamiltonian density is then given by

$$\mathcal{H} = \pi\dot{\phi} + \pi^*\dot{\phi}^* - \mathcal{L}.$$

The covariant expression of charge conservation arises if we define for a space-like surface

$$Q(\sigma) = \int_{\sigma} j_\mu \, d\sigma_\mu. \qquad \ldots\ldots(3.37)$$

From the divergence theorem (2·8) the constancy of Q for different σ immediately follows. It can be verified that for σ given by t = constant (3.37) reduces to the usual definition.

As in the case of neutral fields we now illustrate the general method by its application to a scalar meson field. The Lagrangian is now

$$\mathcal{L} = -\left(\frac{\partial\phi}{\partial x_\nu}\frac{\partial\phi^*}{\partial x_\nu} + \mu^2\,\phi^*\phi\right) \qquad \ldots\ldots(3.38)$$

which in the canonical formalism leads to $\pi = \dot{\phi}^*$, $\pi^* = \dot{\phi}$ and a Hamiltonian

$$\mathcal{H} = \pi^*\pi + \nabla\phi^* \cdot \nabla\phi + \mu^2\,\phi^*\,\phi. \qquad \ldots\ldots(3.39)$$

The covariant commutation relations are again most easily found from the general procedure of § 3.2. This leads immediately to

$$\begin{aligned}[\phi(x)\,,\,\phi(y)] &= 0 = [\phi^*(x)\,,\,\phi^*(y)]\\ [\phi(x)\,,\,\phi^*(y)] &= i\,\Delta(x-y)\end{aligned} \qquad \ldots\ldots(3.40)$$

so that

$$\langle P(\phi(x)\,,\,\phi^*(y)\rangle_0 = i\,\Delta_F(x-y). \qquad \ldots\ldots(3.41)$$

The expansion comparable with (3.19) is

$$\begin{aligned}\phi(x) &= \int\frac{d\mathbf{k}}{(2\pi)^{3/2}}\,\frac{1}{(2\kappa)^{1/2}}\,(a_k e^{ikx} + b_k^*\,e^{-ikx})\\ \phi^*(x) &= \int\frac{d\mathbf{k}}{(2\pi)^{3/2}}\,\frac{1}{(2\kappa)^{1/2}}\,(a_k^*\,e^{-ikx} + b_k\,e^{ikx})\end{aligned} \qquad \ldots\ldots(3.42)$$

with $k_0 = \kappa$, $\kappa = +\sqrt{(\mathbf{k}^2 + \mu^2)}$.
From the covariant commutation relations it follows that a_k, a_k^*, b_k, b_k^* satisfy commutation relations similar to (3.20) of the form

$$[a_k, a_{k'}^*] = \delta(\mathbf{k}-\mathbf{k}') = [b_k, b_{k'}^*] \qquad \ldots\ldots(3.43)$$

which leads to the interpretation of a_k, b_k as annihilation operators, a_k^*, b_k^* as creation operators. With j_μ defined as in (3.35) we find that a_k, a_k^* respectively annihilate and create mesons of charge e and momentum $\mathbf{k}$, energy κ while b_k, b_k^* similarly operate on mesons of charge $-e$.

§ 4. Electromagnetic Field

The electromagnetic field is specified by the scalar and vector potentials ϕ, $\mathbf{A}$ which may be grouped together in a four-vector $A_\mu = (\mathbf{A}, i\phi)$. The observable field strengths $\mathbf{E}$, $\mathbf{H}$ are defined in terms of A_μ by

$$\mathbf{H} = \text{curl}\,\mathbf{A},\ \mathbf{E} = -\dot{\mathbf{A}} - \nabla\phi \qquad \ldots\ldots(4.1)$$

which can be combined in the antisymmetric tensor

$$F_{\mu\nu} = \frac{\partial A_\nu}{\partial x_\mu} - \frac{\partial A_\mu}{\partial x_\nu}. \qquad \ldots\ldots(4.2)$$

The A_μ are not uniquely determined by the field but are arbitrary to a transformation of the kind

$$A_\mu \rightarrow A_\mu + \frac{\partial \chi}{\partial x_\mu} \qquad \ldots\ldots(4.3)$$

called a gauge transformation. The gauge most suited for general treatment is the Lorentz gauge chosen so that $\partial A_\mu/\partial x_\mu = 0$. The field equation satisfied in this case by the components A_μ is

$$\begin{aligned} &\Box A_\mu = 0 \qquad \text{in the absence of sources} \\ \text{or} \quad &\Box A_\mu = -j_\mu \quad \text{in the presence of sources.} \end{aligned} \qquad \ldots\ldots(4.4)$$

In this case the gauge functions χ are limited by the condition $\Box \chi = 0$.

For the most recent developments in radiation theory it is necessary to work in this Lorentz gauge, for appeal has often to be made to relativistic and gauge invariance to eliminate the ambiguities due to the infinities occurring in the theory. However, for some elementary applications the 'Coulomb' gauge can be used, and many difficulties thereby avoided. This consists in a choice such that div $\mathbf{A} = 0$. ϕ then satisfies the equation for an instantaneously propagated static potential, while $\mathbf{A}$ still satisfies the wave equation (4.4) but now only including the transverse part of the source current density. $\mathbf{A}$ is now a transverse field and to each direction of propagation there are only two independent polarizations of light as compared with the four permitted in the Lorentz gauge. The Coulomb gauge is, for example, used by Heitler (1954) in the treatment of all the elementary problems of electrodynamics.

We shall confine the treatment here to a brief discussion of the formalism in the Lorentz gauge. A suitable Lagrangian is that suggested by Fermi (1931)

$$\mathcal{L} = -\tfrac{1}{4}\left(\frac{\partial A_\nu}{\partial x_\mu} - \frac{\partial A_\mu}{\partial x}\right)^2 - \tfrac{1}{2}\left(\frac{\partial A_\mu}{\partial x_\mu}\right)^2. \qquad \ldots\ldots(4.5)$$

This leads to the correct field equations for the A_μ, $\Box A_\mu = 0$. It has to be supplemented by the condition $\partial A_\mu/\partial x_\mu = 0$ in order that Maxwell's equations may hold. The covariant commutation relations derived from equation (4.5) are directly of the form

$$[A_\mu(x)\ ,\ A_\nu(y)] = i\ \delta_{\mu\nu}\ D(x-y) \qquad \ldots\ldots(4.6)$$

where the D-functions are derived from the corresponding Δ-functions by putting $\mu = 0$. The difficulty that the condition $\partial A_\mu/\partial x_\mu = 0$ is in conflict with equation (4.6) is resolved by treating it as a condition limiting possible states.

The field can also be expanded in terms of annihilation and creation operators

$$A_\mu(x) = \int \frac{d\mathbf{k}}{(2\pi)^{3/2}} \frac{1}{(2\kappa)^{1/2}} (a_{k\mu}\, e^{ikx} + a_{k\mu}^*\, e^{-ikx}) \qquad \ldots\ldots(4.7)$$

where $k_0 = \kappa = |\mathbf{k}|$.

The $a_{k\mu}$, $a_{k\mu}^*$ are again annihilation and creation operators for photons of momentum $\mathbf{k}$. There are, however, difficulties associated with the longitudinal and scalar photons contained in equation (4.7), and with the supplementary condition. It was shown by Gupta (1950) and Bleuler (1950) that these difficulties could logically only be avoided by the introduction of an indefinite metric in Hilbert space.

The negative probabilities involved in the use of the indefinite metric only apply to the non-physical longitudinal and scalar photons and so are unobjectionable. We shall not here describe the method of Gupta and Bleuler. In practice it proves possible to ignore the difficulties without error. We conclude here with two vacuum expectation values which will be required in the sequel

$$\langle\{A_\mu(x),\, A_\nu(y)\}\rangle_0 = \delta_{\mu\nu}\, D^{(1)}(x-y) \qquad \ldots\ldots(4.8)$$
$$\langle P(A_\mu(x),\, A_\nu(y))\rangle_0 = \tfrac{1}{2}\,\delta_{\mu\nu}\, D_F\,(x-y).$$

§ 5. Electron-Positron Field

5.1. *Field Equations*

The electron-positron field is described by the wave equation discovered by Dirac which we shall write in the covariant form due to Pauli (1933) as

$$\left(\gamma_\mu \frac{\partial}{\partial x_\mu} + m\right)\psi = 0. \qquad \ldots\ldots(5.1)$$

Here the γ_μ are 4×4 Hermitian matrices satisfying the commutation rules $\gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu = 2\delta_{\mu\nu}$. ψ is a four-component column vector, which we call a spinor. The relativistic invariance of equation (5.1) is not immediately obvious as for the meson and electromagnetic fields. A treatment of this question will be found in the article by Pauli already referred to. It may be noted also that the γ-matrices are related to the original α_k and β matrices of Dirac by the equations

$$\gamma_k = -i\beta\alpha_k, \quad \gamma_4 = \beta. \qquad \ldots\ldots(5.2)$$

The representation most frequently used for α_k, β is

$$\alpha_k = \begin{pmatrix} 0 & \sigma_k \\ \sigma_k & 0 \end{pmatrix}; \quad \beta = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad \ldots\ldots(5.3)$$

where the σ_k are the Pauli spin matrices.

The Hermitian conjugate of ψ, which we denote ψ^* obeys no very simple equation, but if we introduce the adjoint function $\bar{\psi} = \psi^*\gamma_4$ it is found that

$$\frac{\partial}{\partial x_\mu}\bar{\psi}\gamma_\mu - \kappa\,\bar{\psi} = 0. \qquad \ldots\ldots(5.4)$$

The equations (5.1) and (5.4) can be derived from the Lagrangian density

$$\mathcal{L}(x) = -\bar{\psi}(x)\left(\gamma_\mu \frac{\partial}{\partial x_\mu} + m\right)\psi(x). \qquad \ldots\ldots(5.5)$$

That $\mathcal{L}(x)$ is a scalar follows from the covariance properties of the various expressions of the form $\bar{\psi}\,\Gamma\,\psi$, where Γ is some product of γ-matrices. In particular $\bar{\psi}\,\psi$ is a scalar, and $\bar{\psi}\,\gamma_\mu\,\psi$ a four-vector. In the variation of equation (5.5) the four components of ψ, say ψ_ρ, and similarly the four components of $\bar{\psi}$ can all be regarded as independently variable. It is easy then to deduce the field equations (5.1) and (5.4) from the variation principle. The current density vector can also be deduced from equation (5.5); it is

$$j_\mu = ie\,\bar{\psi}\,\gamma_\mu\,\psi. \qquad \ldots\ldots(5.6)$$

We now consider the plane wave solutions of equation (5.1) of the form

$$\psi(x) = u(p)\, e^{ipx}, \quad px = \mathbf{p}.\mathbf{x} - p_0 x_0. \qquad \ldots\ldots(5.7)$$

When the field is quantized these will describe particles of energy p_0, momentum $\mathbf{p}$. The spinors $u(p)$ satisfy the equation

$$(\gamma_\mu p_\mu - imc)\, u(p) = 0. \qquad \ldots\ldots(5.8)$$

Expressions of the sort $\gamma_\mu p_\mu$ occur repeatedly in the Dirac theory and Feynman (1949 a) has introduced the notation $\gamma_\mu p_\mu = \not{p}$. In the older treatments it was usual to consider the set of solutions corresponding to a given value of the momentum $\mathbf{p}$. There are four such solutions, two with $p_0 = +\sqrt{(\mathbf{p}^2 + m^2)}$ the positive energy solutions, and two with $p_0 = -\sqrt{(\mathbf{p}^2 + m^2)}$ the negative energy solutions, the two solutions yielding the two independent spin orientations in each case. Thus for the positive energy solutions we have, using the representation (5.3) for the γ-matrices

$$u^{(1)}(p) = \begin{bmatrix} 1 \\ 0 \\ p_3/(\epsilon + m) \\ (p_1 - ip_2)/(\epsilon + m) \end{bmatrix} \qquad u^{(2)}(p) = \begin{bmatrix} 0 \\ 1 \\ (p_1 + ip_2)/(\epsilon + m) \\ -p_3/(\epsilon + m) \end{bmatrix} \qquad \ldots\ldots(5.9)$$

The solutions have not been normalized, and we define $\epsilon = +\sqrt{(\mathbf{p}^2 + m^2)}$. In recent treatments it has become more usual to group together the positive energy solutions u_+ for momentum $\mathbf{p}$ and the negative energy solutions u_- for momentum $-\mathbf{p}$. The equations defining these solutions are

$$(\not{p} - imc)\, u_+ = 0, \quad (\not{p} + imc)\, u_- = 0 \qquad \ldots\ldots(5.10)$$

where now p_0 has the positive sign $p_0 = \epsilon$. It can be verified immediately that solutions of the form u_+, u_- are always orthogonal to each other, and out of, say, the u_+ an orthogonal pair can always be chosen, e.g. $u^{(1)}$, $u^{(2)}$ above. Thus the four solutions can be chosen to form an orthogonal set.

It is also convenient to make a departure from the older practice regarding normalization in which $u^* u$ is chosen equal to unity, so that in the particle interpretation there is one particle per unit volume. Instead the normalization is chosen so that $u^* u = \epsilon/m$, i.e. one particle per volume m/ϵ. The physical meaning of this normalization is clear when we consider that unit volume in the rest frame of the electron is reduced for a moving electron, owing to the Lorentz–Fitzgerald contraction by a factor $\sqrt{(1 - v^2)} = m/\epsilon$. It is easily verified, with the normalization so chosen, that $\tilde{u}\, u = +1$ for a positive energy state, $\tilde{u}\, u = -1$ for a negative energy state. This result is immediately seen to be true in the rest frame of the electron, and since $\tilde{u}\, u$ is a scalar it holds also for an electron moving with arbitrary momentum. We shall in general denote the set of solutions for a given $\mathbf{p}$ by $u^{(r)}(\mathbf{p})$, $r = 1, 2, 3, 4$, and will observe the convention of using the index i $(=1, 2)$ for a positive energy, j $(=3, 4)$ for a negative energy solution. Introducing the symbol ϵ_r such that $\epsilon_r = +1$, $r = 1, 2$, $\epsilon_r = -1$, $r = 3, 4$ the normalization conditions can be expressed

$$\tilde{u}^{(r)}\, u^{(s)} = \sum_p \tilde{u}_p^{(r)} u_p^{(s)} = \epsilon_r \delta_{rs}. \qquad \ldots\ldots(5.11)$$

The completeness condition then yields that

$$\sum_r \epsilon_r u^{(r)} \bar{u}^{(r)} = \mathbf{1}$$

where $\mathbf{1}$ denotes the unit matrix, or in suffix notation

$$\sum_r \epsilon_r u_\alpha{}^r \bar{u}_\beta{}^r = \delta_{\alpha\beta}. \qquad \ldots\ldots(5.12)$$

5.2 *Projection Operators and Spurs*

The completeness conditions are necessary for the development of a technique for summing over the spins of intermediate states, and averaging over initial and final spins in the evaluation of matrix elements. This technique, using the old normalization, was first introduced by Casimir (1933). Suppose we have to deal with the evaluation of a matrix element of the form

$$M = \sum_i (\bar{v} R w^{(i)})\,(\bar{w}^{(i)} S u) \qquad \ldots\ldots(5.13)$$

where the sum is to be made over the positive energy intermediate states. We introduce the projection operator

$$\Lambda_+(\mathbf{p}) = (p + im)/2m \qquad \ldots\ldots(5.14)$$

with momentum appropriate to the states w. Then $\Lambda_+ w_+ = w_+$ whereas $\Lambda_+ w_- = 0$. M can now be written as a sum over the complete set of intermediate states

$$M = \sum_r (\bar{v} R \epsilon_r \Lambda_+ w^{(r)})\,(\bar{w}^{(r)} S u) = (\bar{v} R \Lambda_+ S u) \qquad \ldots\ldots(5.15)$$

using the completeness relation. In this way any matrix element is reduced to the form $M = (\bar{v} O u)$. What is required is generally $|M|^2$ averaged over the initial and final spins, and this may be evaluated using similar techniques. In fact

$$|M|^2 = M^* M = (u^* O^* \gamma_4 v)\,(\bar{v} O u) = (\bar{u} \gamma_4 O^* \gamma_4 v)\,(\bar{v} O u) \qquad ..(5.16)$$

where O^* denotes the Hermitian conjugate of O. Denoting $\gamma_4 O^* \gamma_4$ by $\bar{O}$ we see that the sum over the initial and final spin states can be written

$$\sum |M|^2 = \sum_r (\bar{u}^{(r)} \bar{O} \Lambda_+{}^f O \epsilon_r \Lambda_+{}^i u^{(r)}) = \sum_r (\bar{u}^{(r)} Q \epsilon_r u^{(r)})$$

with $Q = \bar{O} \Lambda_+{}^f O \Lambda_+{}^i$, and Λ^f, Λ^i the projection operators for the final and initial states. Again using (5.12) this yields

$$\sum |M|^2 = Q_{\rho\rho} = \mathrm{Sp}\, Q.$$

5.3 *Quantization of Electron–Positron Field*

The fields of which the quantization has so far been considered have been Bose fields for which the canonical commutation relationships and Heisenberg equations are consistent with the field equations. For the Dirac field it has been shown by Pauli (1950) and others that the Bose method of quantization leads to physical inconsistencies. The correct method of quantization for the field,

incidentally leading to the exclusion principle was first pointed out by Jordan and Wigner (1928). We deal now essentially with anti-commutation instead of commutation relationships and the fields are expanded in terms of operators a_k satisfying relations of the form

$$a_k\, a_{k'}{}^* + a_{k'}{}^*\, a_k = \{a_k, a_{k'}{}^*\} = \delta_{kk'} \qquad \ldots\ldots(5.17)$$

and

$$\{a_k, a_{k'}\} = 0 = \{a_k{}^*, a_{k'}{}^*\}.$$

For convenience the discrete case is here used as an illustration. Defining, as previously, $N_k = a_k{}^* a_k$ we see immediately that $N_k(1-N_k) = 0$ so that the only possible values of N_k are 0 and 1. a_k and $a_k{}^*$ have again the properties of annihilation and creation operators. The single time commutation relationships leading to equation (5.17) are of the form

$$\{\psi_\rho(\mathbf{x},t),\ \psi_\sigma{}^*(\mathbf{x}',t)\} = \delta_{\rho\sigma}\,\delta(\mathbf{x}-\mathbf{x}') \qquad \ldots\ldots(5.18)$$

together with

$$\{\psi_\rho(\mathbf{x},t),\ \psi_\sigma(\mathbf{x}',t)\} = 0 = \{\psi_\rho{}^*(\mathbf{x},t),\ \psi_\sigma{}^*(\mathbf{x}',t)\}.$$

These relationships, together with the Hamiltonian derived from equation (5.5) and the Heisenberg equations do lead to the correct field equations.

We prefer again to proceed with a direct evaluation of the covariant commutation relations, using the method of § 3.2. The addition to the Lagrangian of an infinitesimal term of the form $\lambda\psi(x)\delta(x-x_1)$, which would lead to Bose commutation rules, is now inappropriate since the spinor quantity is two-valued. One must rather consider additions of the type $\lambda\bar{\theta}\psi(x)\delta(x-x_1)$ where θ is also a spinor quantity. Hence one is led to consider commutators of the form $[\bar{\theta}_1\,\psi(x_1),\ \bar{\psi}(x_2)\,\theta_2]$. In the notation of § 3.2, $\delta_2\psi(x)$ is then determined by the equations

$$\left(\gamma_\mu \frac{\partial}{\partial x_\mu} + m\right)\delta_2\psi = \theta_2\,\delta(x-x_2)$$

$$\delta_2\,\psi = 0, \quad t < t_2.$$

This yields a solution

$$\delta_2\,\psi = -S_R\,(x-x_2)\,\theta_2$$

where the matrix S_R is determined by

$$\left(\gamma_\mu \frac{\partial}{\partial x_\mu} + m\right) S_R(x) = -\,\delta(x),\ S_R(x) = 0,\ t < 0. \qquad \ldots\ldots(5.19)$$

Combining this solution with the similar one for $\delta_1\,\bar{\psi}$ yields the result for the commutator

$$[\bar{\theta}_1\,\psi(x_1),\ \bar{\psi}(x_2)\,\theta_2] = -i\bar{\theta}_1\{S_A(x_1-x_2) - S_R(x_1-x_2)\}\,\theta_2. \qquad \ldots\ldots(5.20)$$

If now $\bar{\theta}_1$, θ_2 are chosen to be spinors anti-commuting with each other, and with all the field quantities the commutator in equation (5.20) can be reduced to $\bar{\theta}_1\{\psi(x_1),\ \bar{\psi}(x_2)\}\,\theta_2$ which leads to the result

$$\{\psi(x_1),\ \bar{\psi}(x_2)\} = -iS\,(x_1-x_2) \qquad \ldots\ldots(5.21)$$

where

$$S(x) = S_A(x) - S_R(x). \qquad \ldots\ldots(5.22)$$

It cannot be pretended that this deduction of the anti-commutation laws for the electron field is a complete one. For a discussion as to how far commutation

rules in general are uniquely determined by the Lagrangian reference may be made to papers by Schwinger (1951a, 1953), also the work of Pauli (1950) referred to above.

The functions $S(x)$ are closely related to the functions $\Delta(x)$, with the appropriate mass m, of § 3.4. Typically we may put

$$S_{\mathrm{R}}(x) = (\gamma_\mu \frac{\partial}{\partial x_\mu} - m)\, \Delta_{\mathrm{R}}(x) \qquad \ldots\ldots(5.23)$$

for then

$$(\gamma_\mu \frac{\partial}{\partial x_\mu} + m)\, S_{\mathrm{R}}(x) = (\Box - m^2)\, \Delta_{\mathrm{R}}(x) = -\delta(x).$$

Integral representations for the S-functions analogous to those already given for the Δ-functions follow immediately from equation (5.23). For example

$$S(x) = i\left(\frac{1}{2\pi}\right)^4 \int_{\mathrm{C}} \frac{p_\mu \gamma_\mu + im}{p^2 + m^2}\, dp \qquad \ldots\ldots(5.24)$$

with C the contour shown in figure 1.
Similarly

$$S_{\mathrm{F}}(x) = -2\left(\frac{1}{2\pi}\right)^4 \int_{\mathrm{C_F}} \frac{p_\mu \gamma_\mu + im}{p^2 + m^2}\, dp. \qquad \ldots\ldots(5.25)$$

5.4. *Annihilation and Creation Operators*

The field operators $\psi(x)$, $\bar{\psi}(x)$ may be expanded in terms of the complete set of plane wave solutions derived in § 5.1 together with operators of annihilation and creation

$$\psi(x) = \int \frac{d\mathbf{p}}{(2\pi)^{3/2}} \left(\frac{m}{\epsilon}\right)^{1/2} \Big[\sum_i a_{pi}\, u^{(i)}(\mathbf{p})\, e^{ipx} + \sum_j b_{pj}^*\, u^{(j)}(\mathbf{p}) e^{-ipx} \Big] \qquad \ldots\ldots(5.26)$$

$$\bar{\psi}(x) = \int \frac{d\mathbf{p}}{(2\pi)^{3/2}} \left(\frac{m}{\epsilon}\right)^{1/2} \Big[\sum_i a_{pi}^*\, \bar{u}^{(i)}(p)\, e^{-ipx} + \sum_j b_{pj}\, \bar{u}^{(j)}(p)\, e^{ipx} \Big]$$

The factor $(m/\epsilon)^{1/2}$ appears because of the method of normalization used for u, and in all cases $px = \mathbf{p} \cdot \mathbf{x} - \epsilon x_0$, with $\epsilon = +\sqrt{(\mathbf{p}^2 + m^2)}$. The significance of the negative energy states was given by Dirac in his 'hole' theory. It will appear immediately that a_{pi} is the operator for the annihilation of a particle of positive energy ϵ and momentum $\mathbf{p}$. The similar annihilation of a particle of negative energy and momentum $-\mathbf{p}$ appears physically as the creation of a positive energy particle of momentum $\mathbf{p}$ (and incidentally with charge opposite to that of the negative energy particle).

Direct application of the covariant commutation rule (5.21) to (5.26) shows that the operators a and b must satisfy the following relations

$$\begin{aligned} \{a_{pi},\, a_{p'i'}^*\} &= \delta(\mathbf{p}-\mathbf{p}')\, \delta_{ii'} \\ \{b_{pj},\, b_{p'j'}^*\} &= \delta(\mathbf{p}-\mathbf{p}')\, \delta_{jj'} \end{aligned} \qquad \ldots\ldots(5.27)$$

which confirms the interpretation of the a, b as annihilation, a^*, b^* as creation operators.

5.5. *Charge Conjugation*

There is a lack of symmetry in the formulation of electron–positron theory given in the previous sections. We start with particles of charge e, say electrons, and the positrons then appear as holes in the negative energy states of the electrons. It is desirable to have a treatment placing both particles on an equal footing, and this can be achieved by means of the charge conjugate field ψ' defined by

$$\psi' = C\,\bar{\psi}^{\mathrm{T}}, \quad \text{with} \quad \gamma_\mu{}^{\mathrm{T}} = -C^{-1}\gamma_\mu\, C. \qquad \ldots\ldots(5.28)$$

Here $\bar{\psi}^{\mathrm{T}}$, $\gamma_\mu{}^{\mathrm{T}}$ are the transposes of $\bar{\psi}$, γ_μ respectively. It is often convenient to ignore the difference between $\bar{\psi}^{\mathrm{T}}$ and $\bar{\psi}$. Then for example the equation (5.4) for $\bar{\psi}$ can be written

$$\left(\gamma_\mu{}^{\mathrm{T}}\frac{\partial}{\partial x_\mu} - m\right)\bar{\psi} = 0.$$

In the presence of an electromagnetic field (cf. § 6.1 below) this becomes

$$\gamma_\mu{}^{\mathrm{T}}\left(\frac{\partial}{\partial x_\mu} + ieA_\mu\right)\bar{\psi} - m\,\bar{\psi} = 0.$$

Introducing ψ' with C defined as in (5.28) leads to

$$\gamma_\mu\left(\frac{\partial}{\partial x_\mu} + ieA_\mu\right)\psi' - m\,\psi' = 0. \qquad \ldots\ldots(5.29)$$

This equation is exactly of the form of that for ψ but with the sign of e changed. So far then as the ψ field describes particles of charge e, the ψ' field describes particles of charge $-e$. The form of the matrix C depends on the reality of the γ-matrix components, and hence on the particular representation of them employed. It can be shown to have the properties (Pauli 1936) $C\,C^* = 1$; $C^{\mathrm{T}} = -C$. With the particular representation (5.3) $C = i\gamma_4\gamma_2$.

The Lagrangian can now be expressed symmetrically between the two fields

$$\mathcal{L} = -\tfrac{1}{2}\bar{\psi}\left(\gamma_\mu\frac{\partial}{\partial x_\mu} + m\right)\psi - \tfrac{1}{2}\bar{\psi}'\left(\gamma_\mu\frac{\partial}{\partial x_\mu} + m\right)\psi'. \qquad \ldots.(5.30)$$

The charge conjugate current density vector is given by

$$j_\mu' = ie\,(\bar{\psi}'\gamma_\mu\,\psi') = ie\,\psi^{\mathrm{T}}\gamma_\mu{}^{\mathrm{T}}\,\bar{\psi}^{\mathrm{T}}. \qquad \ldots\ldots(5.31)$$

In the latter expression the T can safely be dropped from the ψ and $\bar{\psi}$ without risk of confusion. In a classical theory, with ψ, $\bar{\psi}$ commuting, we would have $j_\mu = j_\mu'$, but this no longer holds in the quantum theory. In this case we can re-define $j_\mu(x)$ by the symmetric expression

$$j_\mu = \tfrac{1}{2}\,ie\,(\bar{\psi}\,\gamma_\mu\,\psi - \psi\,\gamma_\mu{}^{\mathrm{T}}\,\bar{\psi}) = -\tfrac{1}{2}\,ie\,[\psi_\alpha, \bar{\psi}_\beta]\,(\boldsymbol{\gamma}_\mu)_{\beta\alpha}. \quad \ldots.(5.32)$$

With this definition the charge conjugate current $j_\mu' = -j_\mu$. The total charge in the field at any time is determined by $Q = -i\int j_4\, d\mathbf{x}$. Using the expression (5.31) for j_4 and the normalization conditions for the $u^{(i)}$, $u^{(j)}$ the value of Q simplifies to

$$Q = e\int d\mathbf{p}\left[\sum_i (a_{pi}^*\, a_{pi} - \tfrac{1}{2}) - \sum_j (b_{pj}^*\, b_{pj} - \tfrac{1}{2})\right]. \qquad \ldots\ldots(5.33)$$

The infinite vacuum expectation value of Q due to the terms $-\frac{1}{2}$ is cancelled out with the present choice of j_μ. We see also, as indicated before, that $a_p^* a_p$ denotes the number of particles of charge e, $b_p^* b_p$ the number of charge $-e$ per unit volume in momentum space. The commutation relations

$$[\bar{\psi}(x), Q] = -e\bar{\psi}(x)\ ;\ [\psi(x), Q] = e\psi(x) \qquad \ldots\ldots(5.34)$$

show also that $\bar{\psi}$ creates charge e or annihilates charge $-e$ and ψ creates charge $-e$ or annihilates charge e.

5.6. *Vacuum Expectation Values*

Finally we deal with some vacuum expectation values for the electron–positron field which will be required in the sequel. The vacuum is defined by $b\Psi_0 = a\Psi_0 = 0$ for all annihilation operators a, b. Then, for example

$$\langle\psi_\alpha(x)\bar{\psi}_\beta(y)\rangle_0 = \Big\langle \int \frac{d\mathbf{p}}{(2\pi)^{3/2}} \left(\frac{m}{\epsilon_p}\right)^{1/2} u_\alpha^i(\mathbf{p})\, e^{ipx} a_{pi} \int \frac{d\mathbf{p}'}{(2\pi)^{3/2}} \left(\frac{m}{\epsilon_{p'}}\right)^{1/2} \bar{u}^{(i')}{}_\beta(\mathbf{p}')\, e^{-ip'y} a_{p'i'}{}^* \Big\rangle_0.$$

Immediately we have

$$\langle a_{pi}\, a_{p'i'}{}^*\rangle_0 = \delta_{ii'}\ \delta(\mathbf{p} - \mathbf{p}')$$

and using the property

$$\sum_i u_\alpha^{(i)}(\mathbf{p})\, u_\beta^{(i)}(\mathbf{p}) = \Lambda_+(\mathbf{p})$$

it follows that

$$\langle\psi_\alpha(x)\ \psi_\beta(y)\rangle_0 = -i\, S_{\alpha\beta}{}^+(x-y) \qquad \ldots\ldots(5.35)$$

where $S^+(x) = (\gamma_\mu\, \partial/\partial x_\mu - m)\ \Delta^+(x)$ with Δ^+ as in (3.26).
Similarly, as in (3.28), it follows that

$$\langle[\psi_\alpha(x), \bar{\psi}_\beta(y)]\rangle_0 = -S_{\alpha\beta}{}^{(1)}(x-y). \qquad \ldots\ldots(5.36)$$

Lastly we consider the chronological products. The Dyson product

$$P(\psi_\alpha(x), \bar{\psi}_\beta(y)) = \psi_\alpha(x)\, \bar{\psi}_\beta(y),\ x_0 > y_0$$
$$= \bar{\psi}_\beta(y)\, \psi_\alpha(x),\ x_0 < y_0$$

so that, with $\bar{S}(x) = -\frac{1}{2}\,\epsilon(x)\, S(x)$

$$\langle\epsilon(x-y)\, P(\psi_\alpha(x), \bar{\psi}_\beta(y))\rangle_0 = -\tfrac{1}{2}S_{\alpha\beta}{}^{(1)}(x-y) + i\,\bar{S}(x-y) = -\tfrac{1}{2}S_F(x-y). \qquad \ldots\ldots(5.37)$$

Another time-ordered product is sometimes introduced, the T-product, defined by $T(U) = (-1)^p\, P(U)$ where U is any series of operators, and p is the number of interchanges of Dirac field operators required to get from the original ordering to that in decreasing time sequence. Equation (5.37) may then be written

$$\langle T(\psi_\alpha(x), \bar{\psi}_\beta(y))\rangle_0 = -\tfrac{1}{2}S_F(x). \qquad \ldots\ldots(5.38)$$

§ 6. Field Interactions

6.1. *Electromagnetic Interactions*

The possible terms that can be inserted in the Lagrangian density to include the interaction between a charged field, specified say by complex field functions

ψ_σ, and the electromagnetic field are limited by several considerations. The new terms must be relativistically invariant and real, and they must comply with the gauge invariance of the electromagnetic field. The condition that there should exist a conserved current density for the non-interacting field is that its Lagrangian should remain invariant to the transformation $\psi_\sigma \to \psi_\sigma e^{i\alpha}$, $\psi_\sigma^* \to \psi_\sigma^* e^{-i\alpha}$. If in this Lagrangian we replace $\partial\psi_\sigma/\partial x_\mu$ where it occurs by $(\partial/\partial x_\mu - ieA_\mu)\psi_\sigma$ and similarly $\partial\psi_\sigma^*/\partial x_\mu$ by $(\partial/\partial x_\mu + ieA_\mu)\psi_\sigma^*$ then the new Lagrangian is invariant to the transformation

$$\psi_\sigma \to \psi_\sigma e^{i\alpha}, \quad \psi_\sigma^* \to \psi_\sigma^* e^{-i\alpha} \qquad \ldots\ldots(6.1)$$

$$A_\mu \to A_\mu + \frac{1}{e}\frac{\partial \alpha}{\partial x_\mu}.$$

This method of introducing the electromagnetic interaction satisfies all the necessary conditions, and is indeed uniquely determined by them.

In the case of the scalar meson field in interaction with the electromagnetic field, writing the complete Lagrangian in the form $\mathcal{L} = \mathcal{L}_0 + \mathcal{L}^{\mathrm{I}}$ where $\mathcal{L}_0$ gives the sum of the Lagrangians for the two free fields then

$$\mathcal{L}^{\mathrm{I}} = -ieA_\mu\left(\phi^*\frac{\partial\phi}{\partial x_\mu} - \frac{\partial\phi^*}{\partial x_\mu}\phi\right) + e^2A_\mu^2\phi\phi^*. \qquad \ldots\ldots(6.2)$$

A similar evaluation for the electron–positron field, with the Lagrangian in its charge conjugate form

$$\mathcal{L} = -\tfrac{1}{2}\left\{\bar{\psi}\left(\gamma_\mu\frac{\partial}{\partial x_\mu} + m\right)\psi + \psi\left(\gamma_\mu{}^{\mathrm{T}}\frac{\partial}{\partial x_\mu} - m\right)\bar{\psi}\right\}$$

yields

$$\mathcal{L}^{\mathrm{I}} = \tfrac{1}{2}ie[\bar{\psi}\gamma_\mu A_\mu\psi - \psi\gamma_\mu{}^{\mathrm{T}}A_\mu\bar{\psi}]. \qquad \ldots\ldots(6.3)$$

The coupled field equations in this case have the form

$$\left(\gamma_\mu\left(\frac{\partial}{\partial x_\mu} - ieA_\mu\right) + m\right)\psi = 0, \quad \Box A_\mu = -\tfrac{1}{2}ie\left\{\bar{\psi}\gamma_\mu\psi - \psi\gamma_\mu{}^{\mathrm{T}}\bar{\psi}\right\} \quad \ldots(6.4)$$

together with the supplementary condition $(\partial A_\mu/\partial x_\mu)\bar{\psi} = 0$.

6.2. *Other Interactions*

Electromagnetic interactions are alone in their unique specification by requirements of gauge invariance. However, requirements of reality and relativistic invariance remain, and if to these are added the postulates, in analogy with the electromagnetic field case, that the interactions are to be local, linear in the fields, and not involving derivatives of the Fermi field functions, the arbitrary element is largely removed. For example the interaction between a nucleon field ψ and a pseudoscalar meson ϕ must be of one of the two forms

$$\mathcal{L}^{\mathrm{I}} = iG\bar{\psi}\gamma_5\psi\phi \quad \text{or} \quad \mathcal{L}^{\mathrm{I}} = g\bar{\psi}\gamma_5\gamma_\mu\psi\frac{\partial\phi}{\partial x_\mu}. \qquad \ldots\ldots(6.5)$$

The second type of coupling is often referred to as a derivative coupling, and we shall see later that it leads to a non-renormalizable theory. The interaction

between π-mesons and nucleons is now believed to be of the first type. However, it must be emphasized how much weaker our belief in this meson coupling must be as compared with the electromagnetic coupling. There is no *a priori* reason against the occurrence of non-linear couplings in nature. Only a desire for simplicity restricts us to the linear ones, which will at least form a valid approximation in some energy regions. Unfortunately the detailed calculations which would form the justification for the use of the equation (6.5) are not at present sufficiently reliable to permit definite conclusions being drawn.

§ 7. Solution of the Equations of Field Theory

7.1. *Schrödinger Representation*

In the previous sections the equations of motion of coupled quantum fields have been developed in the Heisenberg representation. We have seen that, in this representation, there result coupled partial differential equations for the field operators. While the commutation relations at a given time are still the canonical ones they cannot be generalized, as for the bare fields, to arbitrary world points. Also we can no longer make an expansion of the type (5.26) in terms of a complete set of wave functions and creation and annihilation operators. The older method of formulating the problem was to use the Schrödinger representation. In it the field operators are no longer time dependent and may be equated to the Heisenberg operators at some particular time, say for convenience $t = 0$. Formally the relationship between the two representations is as defined below: In the Heisenberg representation

$$\Psi_{\mathrm{H}} = \text{const.}, \qquad i\frac{dO_{\mathrm{H}}}{dt} = [O_{\mathrm{H}}(t), H]. \qquad \ldots\ldots(7.1)$$

In the Schrödinger representation

$$i\frac{\partial \Psi_{\mathrm{S}}(t)}{\partial t} = H\,\Psi_{\mathrm{S}}(t), \qquad O_{\mathrm{S}} = \text{const.} \qquad \ldots\ldots(7.2)$$

so that

$$\Psi_{\mathrm{S}}(t) = T(t, 0)\,\Psi_{\mathrm{S}}(0), \qquad O_{\mathrm{H}}(t) = T^{-1}(t,0)\,O_{\mathrm{S}}\,T(t,0) \quad \ldots\ldots(7.3)$$

where $i\dfrac{\partial}{\partial t}T(t, 0) = H\,T(t, 0), \qquad T(0, 0) = 1.$

This may be solved to yield $T = \exp(-iHt)$. The Hamiltonian H is a constant of the motion, and for it no distinction need be made between the two representations.

In the Schrödinger picture the Hamiltonian density is expressed in terms of the field functions $\psi(\mathbf{x})$ and their conjugate momenta $\pi(\mathbf{x})$. These satisfy the canonical commutation relations and so can be expanded just like the bare particle operators in terms of creation and annihilation operators. The Schrödinger equation for the state vector of the system takes the form

$$i\frac{\partial \Psi}{\partial t} = (H_0 + H^1)\,\Psi \qquad \ldots\ldots(7.4)$$

where H_0 is the sum of the Hamiltonians for the interacting fields, and H^1 is that

for the interaction. The solution of equation (7.4) may then be attempted by the methods of perturbation theory.

We illustrate the formalism by two examples:

(i) *Interaction of the electron–positron with the electromagnetic field.*

The Lagrangian density is $\mathcal{L} = \mathcal{L}_0 + \mathcal{L}^1$ where $\mathcal{L}_0$ is the sum of the densities for the free electron and electromagnetic fields, and

$$\mathcal{L}^1 = -j_\mu A_\mu \; ; \; j_\mu = \tfrac{1}{2}ie\,(\bar{\psi}\gamma\,\psi - \bar{\psi}'\gamma_\mu\,\psi').$$

The conjugate momenta are defined by the relation $\pi = \partial\mathcal{L}/\partial\dot{\phi}$. For this case $\mathcal{L}^1$ contains no derivatives of the field functions. So the conjugate momenta are exactly as for the bare fields. The Hamiltonian density then has the form

$$\mathcal{H} = \mathcal{H}_0 + \mathcal{H}^1 \quad \text{with} \quad \mathcal{H}^1 = -\mathcal{L}^1. \qquad \ldots\ldots(7.5)$$

In this case it may be noted that the interaction Hamiltonian density is a scalar.

(ii) *Interaction of charged scalar mesons with electromagnetic field.*

The Lagrangian for the interacting fields is

$$\mathcal{L} = -\left\{\left(\frac{\partial}{\partial x_\nu} + ieA_\nu\right)\phi^*\right\}\left\{\left(\frac{\partial}{\partial x_\nu} - ieA_\nu\right)\phi\right\} - \mu^2\phi^*\phi + \mathcal{L}_e$$

where $\mathcal{L}_e$ is the Lagrangian density of the bare electromagnetic field. From this we have

$$\mathcal{L}^1 = -ieA_\nu\left(\phi^*\frac{\partial\phi}{\partial x_\nu} - \frac{\partial\phi^*}{\partial x_\nu}\phi\right) - e^2A_\nu^2\phi^*\phi \qquad \ldots\ldots(7.6)$$

and the conjugate momenta for the meson field are changed from those of the bare fields e.g. $\pi = \partial\phi^*/\partial t - ieA_0\,\phi^*$. The interaction Hamiltonian density is no longer equal to $-\mathcal{L}^1$ but appears as

$$\mathcal{H}^1 = -ieA_0(\pi\phi - \pi^*\phi^*) - ie\mathbf{A}.\;((\nabla\phi^*)\,\phi - (\nabla\phi)\,\phi^*) + e^2\mathbf{A}^2\phi^*\phi. \qquad \ldots\ldots(7.7)$$

The new π together with ϕ obeys the canonical commutation relations and can be expanded identically with the bare field Heisenberg operators at a given time.

7.2. *Interaction Representation*

An alternative method of dealing with the equations was recently developed in a series of papers by Tomonaga and collaborators (1946, 1947, 1948), Schwinger (1948) and others. It bears a relationship to an earlier treatment of Dirac, Fock and Podolsky. The method is to introduce a new representation intermediate between the Heisenberg and Schrödinger representations such that in it the field operators satisfy the bare field equations of motion. It is then possible to evaluate commutators of field quantities at arbitrary world points, while at the same time expanding these quantities in bare field annihilation and creation operators. The method has the advantage over the Schrödinger representation that it can be expressed in a covariant manner throughout. This is important when dealing with the divergent expressions arising from such phenomena as the polarization of the vacuum.

We start from the Heisenberg representation in which the Hamiltonian H,

itself time independent, is supposed divided into two parts, which will each separately be time dependent

$$H = H_0(t) + H^1(t) \qquad \ldots\ldots(7.8)$$

where $H_0(t)$, $H^1(t)$ are respectively the bare field and interaction parts of the total Hamiltonian. Hamiltonians will be assumed to be in the Heisenberg representation unless it is indicated otherwise explicitly in the notation. Then if $O(t)$ is any field operator the corresponding interaction representation operator $O_i(t)$ is defined by

$$O_i(t) = \exp(iH_0(0)t)\, O_H(0) \exp(-iH_0(0)t). \qquad \ldots\ldots(7.9)$$

$O_H(0)$ may be taken as the Schrödinger representation operator O_S. Equation (7.9) may be written as

$$O_i(t) = U(t,0)\, O_H(t)\, U^{-1}(t,0) \qquad \ldots\ldots(7.10)$$

where $U(t,0) = \exp(iH_0(0)\, t) \exp(-iHt)$. From this definition it is easy to verify that $U(t,0)$ satisfies the differential equation

$$i\frac{\partial}{\partial t} U(t,0) = H_i^1(t)\, U(t,0), \text{ with } U(0,0) = 1. \qquad \ldots\ldots(7.11)$$

Also, if we denote by $\Psi_i(t)$ the state vector in the interaction representation then, since $\Psi_i(t) = U(t,0)\, \Psi_H$ it follows that

$$i\frac{\partial \Psi_i}{\partial t} = H_i^1(t)\, \Psi_i. \qquad \ldots\ldots(7.12)$$

The definition (7.9) shows that the interaction representation field operators do indeed have the time variation of the corresponding bare field operators, with which they can be identified. In particular, they will obey the commutation relations for such operators. Equation (7.12) thus gives the time variation of the interaction representation state vector as governed by an interaction Hamiltonian expressed in terms of bare field quantities. These latter can be replaced by the annihilation and creation operator expansions, for example, of equation (5.26).

As an illustration of the application of the method we take the two cases for which the Schrödinger formalism has already been given.

(i) *Electron–electromagnetic field interaction.*

In this case no difficulties arise, as $\mathcal{L}^1$ contains no derivatives and the canonical conjugate momenta have the same form for interacting as for bare fields. Directly from equation (7.5) we have

$$\mathcal{H}_i^1(x) = -j_\mu(x)\, A_\mu(x) \qquad \ldots\ldots(7.13)$$

where now A_μ and the operators involved in the definition (5.32) of j_μ are in the interaction representation.

(ii) *Meson–electromagnetic field interaction.*

The Schrödinger interaction Hamiltonian is given in equation (7.7) in terms of the field operators and their conjugates. In accordance with equation (7.9) the interaction representation form, $\mathcal{H}_i^1(x)$ is found by giving the field operators in (7.7) the form appropriate to bare fields. The equation earlier given relating

π, ϕ no longer holds. Rather we have $\pi_i(x) = \dot{\phi}_i^*(x)$ and similarly for π^*. This leads to the expression for $\mathcal{H}_i^{\,1}(x)$ in terms of the field operators and their derivatives, after the elimination of the π's as

$$\mathcal{H}_i^{\,1}(x) = ieA_\nu \left(\phi^* \frac{\partial \phi}{\partial x_\nu} - \frac{\partial \phi^*}{\partial x_\nu} \phi \right) + e^2 A_\nu^2 \phi^* \phi + e^2 A_0^2 \phi^* \phi. \quad \ldots(7.14)$$

It should be noted that this is no longer relativistically invariant. If the canonical formalism is developed on a space-like surface, normal n_μ, as described in § 2.2, then the last term of $\mathcal{H}_i^{\,1}(x)$ must be written $e^2(n_\mu A_\mu)^2 \phi^* \phi$. It has been shown by Matthews (1949) that the contribution to the S-matrix from such normal dependent terms is completely cancelled out.

A further generalization to the formalism of the interaction representation can be made when the space-like surfaces considered in § 2.2 are no longer required to be plane. The state of the system is then defined by interaction representation state vectors $\Psi(\sigma)$, functionals of the surface σ. In the notation of Schwinger (1948) the equation of motion for $\Psi_i(\sigma)$, replacing (7.12) is of the form

$$i \frac{\delta \Psi_i(\sigma)}{\delta \sigma(x)} = \mathcal{H}_1^{\,1} \Psi_i(\sigma) \qquad \ldots\ldots(7.15)$$

where the functional derivative is defined by

$$\frac{\delta \Psi(\sigma)}{\delta \sigma(x)} = \operatorname*{Lim}_{\delta\omega \to 0} \frac{\Psi(\sigma') - \Psi(\sigma)}{\delta \omega} \qquad \ldots\ldots(7.16)$$

where $\delta\omega$ is the real space time element enclosed between the neighbouring surfaces σ' and σ which differ only in the neighbourhood of a world point x. We shall not have occasion to make use of this generalization.

7.3. *Dyson's Solution for the S-Matrix*

We turn now to the solution given by Dyson of the equation for the operator U leading from the Heisenberg to the interaction representation. The arbitrary initial time at which the two representations coincide will now be taken as $t = t_0$ so that the equation for U is

$$i \frac{\partial}{\partial t} U(t, t_0) = H_i^{\,1}(t)\, U(t, t_0) \qquad \ldots\ldots(7.17)$$

where $$U(t_0, t_0) = 1.$$

The operator U describes the development of the interaction representation state vector $\Psi_i(t) = U(t, t_0)\, \Psi_i(t_0)$.

In particular if we take the limit of U as $t \to \infty$, $t_0 \to -\infty$ we obtain the S-matrix, $S = U(\infty, -\infty)$ with the property

$$\Psi_i(\infty) = S\, \Psi_i(-\infty). \qquad \ldots\ldots(7.18)$$

The differential equation (7.17) can be replaced by an integral equation which includes the boundary condition. Defining $U(t) = U(t, -\infty)$ this is

$$U(t) = 1 + (-i) \int_{-\infty}^{t} H_i^{\,1}(t')\, U(t')\, dt'. \qquad \ldots\ldots(7.19)$$

An iterated solution can be given of (7.19) in the form

$$U(t) = 1 + (-i)\int_{-\infty}^{t} H_i^I(t_1)\,dt_1 + (-i)^2\int_{-\infty}^{t} H_i^I(t_1)\,dt_1\int_{-\infty}^{t_1} H_i^I(t_2)\,dt_2. + \ldots \quad \ldots(7.20)$$

It was remarked by Dyson that the nth power term in this series can be expressed in the form

$$\frac{(-i)^n}{n!}\int_{-\infty}^{t}\int_{-\infty}^{t}\ldots\int_{-\infty}^{t} dt_1 \ldots dt_n\, P(H_i^I(t_1), H_i^I(t_2) \ldots H_i^I(t_n)) \quad \ldots\ldots(7.21)$$

where the chronological product P has already been defined as arranging the operators in increasing time sequence from right to left. This result follows from the fact that the integrand of (7.21) is symmetric in the $t_1 \ldots t_n$ and hence must be $n!$ times the value found when $t_1 \ldots t_n$ in it are selected in an arbitrary order. This is just the form found in the series for $U(t)$, the order there being $t_n < t_{n-1} < \ldots < t_1$. Introducing the Hamiltonian densities and taking the limit as $t \to \infty$ we obtain the Dyson formula for the S-matrix

$$S = \sum_n \frac{(-i)^n}{n!}\int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 \ldots \int_{-\infty}^{\infty} dx_n\, P(\mathcal{H}_i^I(x_1) \ldots \mathcal{H}_i^I(x_n)) \ldots\ldots(7.22)$$

where now dx_n stands for $dt_n\, d\mathbf{x}_n$. S gives the overall change in the system due to the interaction, and it can be used, for example for the evaluation of cross sections. Suppose the state of the system at $t = -\infty$ can be represented as a set of 'bare' particles of momentum $\mathbf{p}_1, \mathbf{p}_2 \ldots \mathbf{p}_n$. Then the probability amplitude that at $t = +\infty$ the system consists of bare particles $\mathbf{p}_1', \mathbf{p}_2' \ldots$ is just the matrix element of S, $(\mathbf{p}' \mid S \mid \mathbf{p})$. We shall see that the fact that actual collisions are not between 'bare' particles, but rather particles surrounded by their self-field, is not an important distinction, being completely covered by the renormalization of mass and charge.

7.4. *Evaluation of S-matrix Elements*

An elegant formalism has been found for the evalution of S-matrix elements. This was originally developed by Feynman (1949 a, b) who described the propagation of an electron in space–time under the influence of an external field, or of other electrons, with the help of a graphical notation. Dyson (1949 a, b) showed the equivalence of the Feynman method to the normal techniques of field quantization described above. Later Wick (1950) demonstrated how the Feynman graphs could be extracted from the formalism of the S-matrix by purely algebraic means. Wick's treatment involves the definition of a new ordered form for the product of a number of creation and annihilation operators, the so-called normal product defined by

$$N(UV \ldots W) = (-1)^p\, U'\, V' \ldots W' \quad \ldots\ldots(7.23)$$

where $U', V' \ldots W'$ denote the same operators as $U, V \ldots W$ but ordered so that all the annihilation operators are to the right and creation operators to the left, and p is equal to the number of interchanges of Dirac operators required to go from the original ordering to the normal ordering. Products of this sort were used previously by Stueckelberg (1944) and Houriet and Kind (1949).

We consider the nth order term in the operator S, S_n say, and for the present confine attention to electron–positron theory when

$$\mathcal{H}_i^{\mathrm{I}}(x) = -\tfrac{1}{2}ie\,[\bar{\psi}_\beta, \psi_\alpha]\,(\gamma_\mu)_{\beta\alpha}A_\mu.$$

It will be seen later that for most purposes the charge conjugate expression for the current density can be ignored. Its only function, as we saw in § 5.5, is to reduce to zero the vacuum expectation value of the current. Accordingly, for a typical term in the integrand of S_n we have, ignoring suffixes and numerical constants,

$$T(X_1, X_2, \ldots X_n), \text{ where } X_1 = \bar{\psi}(x_1)\,A(x_1)\,\psi(x_1) \text{ etc.}$$

Here it should be noted that the T product (cf. (5.38)) may be used in place of the P product since the Dirac operators in the X_i occur in pairs. Wick has given a theorem for the reduction of such a product to the sum of its normal constituents

$$T = \sum_c T_c \qquad \ldots\ldots(7.24)$$

where a typical normal constituent is defined by

$$T_c = (-1)^p\,(T(UV))_0\,(T(U'V'))_0 \ldots N(AB \ldots D). \qquad \ldots.(7.25)$$

Here U, V, U', $V' \ldots A$, $B \ldots D$ are each single field operators of the electron or electromagnetic field, and their product represents an ordering of the product $X_1 \ldots X_n$. p is again the number of interchanges of Dirac operators in going from T to T_c. The sum in equation (7.24) is over all possible constituents. It is clear that, in the pairing of operators U, V to make up constituents, U, V must either both be electron or photon operators. Also from § 5.5 we know that products of the type $T(\psi, \psi)$ have zero vacuum expectation value, so that for the electron field only pairings of the type $T(\psi, \bar{\psi})$ need be considered. Incidentally, at this point the only effect of using the full current operator (5.32) becomes clear. Normal constituents of T occur in which $\bar{\psi}(x_i)$, say, is paired with $\psi(x_i)$. When we take the vacuum expectation value of such a pairing we get a component of the vacuum current which should be zero, and is so with the full definition of (5.32). Hence such pairings must be ignored; subject to this proviso j_μ can then be used in its simple form.

With these restrictions the normal constituent T_c may be associated with a Feynman diagram as follows. We have n points $x_1, x_2 \ldots x_n$. For each associated pair of Dirac operators $\psi(y)\bar{\psi}(x)$ we draw a directed (electron) line from x to y. For each pair of photon operators $\phi(x)$, $\phi(y)$ we draw a dotted (photon) undirected line from x to y. For each unpaired $\bar{\psi}(x)$, $\psi(x)$ we draw a line from x to the edge of the diagram, directed respectively out from x in the first case and in to x in the second. For an unpaired $\phi(x)$ we draw a dotted line to the edge of the diagram. Physically the unpaired operators annihilate the incoming and create the outgoing particles. The paired operators create and annihilate virtual particles. In passing we may note that a positron created at x and annihilated at y is according to the above rules denoted by a line drawn from y to x.

In considering any simple process, for instance, Compton scattering, in a low order it is generally easier to start from the graphs which can be drawn for the process and find the contributions to the S-matrix from these rather than algebraically to write down the normal constituents. For Compton scattering in

second (lowest) order for example the possible graphs are those shown in figure 2.

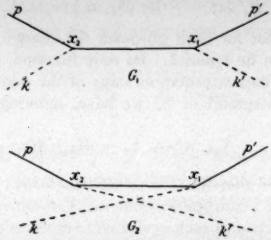


Figure 2. Lowest order graphs for Compton scattering.

The method of finding the S-matrix contribution from a given graph is now clear from the preceding analysis. If we still confine attention to the interaction of electrons and the electromagnetic field, then for a graph with n vertices we are interested in the term of S, S_n say,

$$S_n = \frac{(-1)^n e^n}{n!} \int_{-\infty}^{\infty} dx_1 \ldots \int_{-\infty}^{\infty} dx_n \; T(\bar{\psi}\gamma_\mu \psi A_\mu(x_1) \ldots \bar{\psi}\gamma_\mu \psi A_\mu(x_n)). \quad \ldots(7.26)$$

The graph corresponds with a definite normal constituent of T and its contribution to S_n is of the form

$$S_n(G) = \frac{(-1)^n e^n}{n!} (-1)^p \int_{-\infty}^{\infty} dx_1 \ldots \int_{-\infty}^{\infty} dx_n \, \langle T(\psi_\alpha(x), \bar{\psi}_\beta(y)) \rangle_0 \ldots$$
$$\times \langle T(A_\mu(x'), A_\nu(y')) \rangle_0 \ldots N \quad \ldots\ldots(7.27)$$

where N is the normal product for the creation and annihilation of external particles, and p is the number of interchanges of Dirac operators between the ordering of S_n and of $S_n(G)$. Then using equations (3.32), (5.38) we may replace

$$\langle T(\psi_\alpha(x), \bar{\psi}_\beta(y)) \rangle_0 = -\tfrac{1}{2} S_{F\alpha\beta}(x-y)$$
$$\langle T(A_\mu(x'), A_\nu(y') \rangle_0 = \tfrac{1}{2} \delta_{\mu\nu} D_F(x-y)$$

where for completeness here we recall the definitions

$$\tfrac{1}{2}\Delta_F(x) = \frac{-i}{(2\pi)^4} \int_{C_F} \frac{e^{ikx}}{k^2+\mu^2} dk \; ; \; \tfrac{1}{2} S_F(x) = \left(\frac{1}{2\pi}\right)^4 \int_{C_F} \frac{\gamma p + im}{p^2 + m^2} e^{ipx} dp \quad \ldots(7.28)$$

$D_F(x)$ is found by putting $\mu = 0$ in $\Delta_F(x)$.

In the construction of a graph there is generally a quite arbitrary numbering of the vertices $1 \ldots\ldots n$. Summing over all the permutations of these vertices then removes the factor $1/n!$ from $S_n(G)$. As a simple example of the procedure we take the graph G_1 for Compton scattering. For this graph

$$S_2(G_1) = \frac{e^2}{2} \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 \, \langle T(\psi_{\alpha'}(x_1), \bar{\psi}_\beta(x_2)) \rangle_0 \, \bar{\psi}_\alpha(x_1) (\gamma_\mu)_{\alpha\alpha'} \psi_{\beta'}(x_2) (\gamma_\nu)_{\beta\beta'}$$
$$\times A_\mu(x_1) A_\nu(x_2). \quad \ldots\ldots(7.29)$$

There is an exactly equivalent graph with x_1, x_2 interchanged. Also it is much more convenient to preserve the order of the electron field operators so that products can be expressed in a matrix rather than a suffix notation. Summing for the two graphs we have

$$S_2(G_1) = e^2 \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2\, \bar{\psi}(x_1)\, \gamma_\mu \{ -\tfrac{1}{2} S_F(x_1 - x_2) \}\, \gamma_\nu\, \psi(x_2)\, A_\mu(x_1)\, A_\nu(x_2). \quad \ldots\ldots (7.30)$$

It should be noted that equation (7.30) also contains the contribution of G_2 (figure 2). For the evaluation of transition probabilities the matrix element of S is required between the sets of bare particle states under consideration. In the case of Compton scattering, for example, in the notation of figure 2 we require $(\mathbf{p}', \mathbf{k}' \mid S \mid \mathbf{p}, \mathbf{k})$. This can be found using the Fourier integral expansions already given for the various field operators. For example, according to equation (5.26), the electron annihilation part of $\psi(x_2)$ required here is given in the form

$$\psi(x_2) = \int d\mathbf{q}\, a(\mathbf{q})\, f(x_2, \mathbf{q})$$

where f is a c-number. This has to operate on the one electron state $|\mathbf{p}\rangle$, annihilating the electron. Now $|\mathbf{p}\rangle$ can be written $a_p^* |0\rangle$ where $|0\rangle$ is the vacuum state. Thus using the anti-commutation relations and the vacuum properties

$$\psi(x_2)\, |\mathbf{p}\rangle = \int d\mathbf{q}\, f(x_2, \mathbf{q}) \{ a(q), a^*(p) \} |0\rangle = f(x_2, \mathbf{p}) |0\rangle. \quad \ldots\ldots (7.31)$$

In this way all the operators in $S(G)$ are used up. In the case of Compton scattering, for graph G_1,

$$M_1 = (\mathbf{p}', \mathbf{k}' | S_2 | \mathbf{p}\, \mathbf{k}) = e^2 \int dx_1 \int dx_2 \left(\frac{m}{\epsilon'}\right)^{1/2} \frac{\bar{u}^{i'}(p')}{(2\pi)^{3/2}} \exp(-ip'x_1).\gamma_\mu \left(-\frac{1}{(2\pi)^4}\right)$$

$$\int \exp\{ iq(x_1 - x_2) \} S_F(q)\, dq\, \gamma_\nu \left(\frac{m}{\epsilon}\right)^{1/2} u^i(p) \exp(ipx_2) \frac{e'_\mu \exp(-ik'x_1)}{(2\pi)^{3/2}(2\kappa')^{1/2}} \frac{e_\nu \exp(ikx_2)}{(2\pi)^{3/2}(2\kappa)^{1/2}}. \quad \ldots\ldots (7.32)$$

Here i, i' are the initial and final electron spin states, and e, e' are the initial and final photon polarization vectors. ϵ, ϵ' are both positive electron energies and

$$S_F(q) = (q_\mu \gamma_\mu - im)^{-1}. \quad \ldots\ldots (7.33)$$

The space integrals can immediately be carried out, yielding energy momentum conservation at each vertex. The integral over q converts this into overall energy momentum conservation reducing the expression for M_1 to

$$M_1 = (2\pi)^4 \delta(p + k - p' - k') \left\{ \frac{1}{(2\pi)^{3/2}} \right\}^4 (-e^2) \left\{ \frac{m}{\epsilon'} \frac{m}{\epsilon} \frac{1}{2\kappa'} \frac{1}{2\kappa} \right\}^{1/2} \quad \ldots\ldots (7.34)$$

$$\times\, \bar{u}^{i'}(p')\, e_\mu' \gamma_\mu S_F(p + k)\, e_\nu \gamma_\nu u^i(p).$$

In general for more complicated processes matrix elements similar to M_1 will be found. The occurrence of the various factors in equation (7.34) has a simple physical interpretation and the form (7.34) can thus be generalized to other cases without a detailed derivation. Thus there is a residual factor $(2\pi)^4$ left over from the integration over space time of the vertex points after the compensating

factors $(2\pi)^{-4}$ from the electron (or photon) propagation functions $S_F(D_F)$ have been included. The δ-function expresses overall momentum conservation, and $-e^2$ corresponds to two interactions with the electromagnetic field. The other factors outside the spinor matrix element are associated with the choice of normalization—$(1/2\pi)^{3/2}$ for each 'external' particle, incoming or outgoing, $(m/\epsilon)^{1/2}$ for each external electron, $(1/2\kappa)^{1/2}$ for external quantum. The spinor matrix element between the final and initial electron states is of an operator made by taking in order $e_\mu\gamma_\mu$ for a vertex at which a photon, polarization vector e_μ interacts with the electron, $S_F(p)$ for an internal (virtual) electron propagating with energy-momentum p and $D_F(p)$ for a similar photon.

7.5. *Evaluation of Cross Sections*

We have seen that the S-matrix element between two bare particle states, containing electrons and photons, N incoming, N' outgoing particles can be expressed in the form

$$(\mathbf{p}'|S|\mathbf{p})=(2\pi)^4\delta_4(P-P')\left(\frac{1}{(2\pi)^{3/2}}\right)^{N+N'} zz'M \qquad \ldots\ldots(7.35)$$

where $\mathbf{p}$, $\mathbf{p}'$ are respectively the momenta of the incoming and outgoing particles and $P=\Sigma p$, $P'=\Sigma p'$. z,z' are here used to denote the external particle normalization factors, $(m/\epsilon)^{1/2}$ for each electron, $(1/2\kappa)^{1/2}$ for each photon, and M is the spinor matrix element. From a knowledge of S we wish to derive the cross section for the scattering process concerned, which may, of course, include particle creation. The scattering cross section is only defined for two incident particles, so, strictly, we may directly put $N=2$. However, the transition probability from which the scattering cross section is derived applies to the general case.

With the normalization of states we have chosen

$$dW=|(\mathbf{p}' \mid S \mid \mathbf{p})|^2 d\mathbf{p}';\ d\mathbf{p}'=d\mathbf{p}_1'\, d\mathbf{p}_2' \ldots d\mathbf{p}_{N'}' \qquad \ldots\ldots(7.36)$$

represents the overall transition probability from the given initial state to a state in which the N' final particles have momentum in the momentum space volume element $d\mathbf{p}'$. This is directly a meaningless quantity involving as it does the square of a δ-function. The reason is that in equation (7.36) the probability of transition is given for an infinite space after an infinite time. The physically significant quantity is the transition probability W' per unit volume in space, per unit time, when the incoming particles are normalized to one per unit volume. To give an adequately rigorous derivation of W' it is desirable to enclose the whole system in a normalization volume V, making the appropriate change to a discrete expansion of the field operators (cf. Pauli 1951). We shall not consider this in detail but merely give a treatment of equation (7.36) which suggests the correct solution. $\delta(E)$ can be defined by

$$\delta(E)=\operatorname{Lim}_{T\to\infty}\frac{1}{2\pi}\,\frac{\sin^2\frac{1}{2}ET}{(\frac{1}{2}E)^2T}\,.$$

In place thus of $(\delta(E))^2=\delta(E)\delta(0)$ we are led to put $\lim_{T\to\infty}[(T/2\pi)\delta(E)]$; as it were one of the energy δ-functions is spread out over an energy interval $1/T$ as it is if the system is only under observation for a time T. Analogous ideas apply to the momentum δ-functions, and we are led to the expression

$$dW=(2\pi)^8\left(\frac{1}{2\pi}\right)^{3(N+N')}\delta_4(P-P')\,\frac{TV}{(2\pi)^4}\,z^2z'^2\,|M|^2\,d\mathbf{p}'\quad\ldots\ldots(7.37)$$

as the transition probability for a space time volume TV. The particle states used in equation (7.37) are normalized to a density $(1/2\pi)^{3/2}$; hence for W' defined above

$$dW'=(2\pi)^4z^2z'^2\,|M|^2\,\delta_4(P-P')\,\frac{d\mathbf{p}_1'}{(2\pi)^3}\quad\ldots\quad\frac{d\mathbf{p}_{N'}}{(2\pi)^3}.\quad\ldots\ldots(7.38)$$

The corresponding cross section $d\sigma$ is obtained by dividing dW' by $|\mathbf{v}_1-\mathbf{v}_2|$ where $\mathbf{v}_1$, $\mathbf{v}_2$ are the velocities of the two incident particles, supposed in the same direction. The fact that the resulting total cross section is relativistically invariant is not immediately obvious from equation (7.38). However, by absorbing some of the external normalization factors it is possible to rewrite (7.38) in a manifestly invariant form. A typical $d\mathbf{p}'$ can be replaced by $2\epsilon' dp'\delta(p'^2-m^2)$ where as usual $dp'=d\mathbf{p}'dp_0'$. Since $\delta(p'^2-m^2)=(1/2\epsilon')\{\delta(p_0'-\epsilon')+\delta(p_0'+\epsilon')\}$ this is valid provided it is understood that integration is to be carried out only over that part of momentum space for which $p_0'>0$. If a photon is involved the same formulae are valid with $m=0$. dW' can thus be written in the form

$$dW=(2\pi)^4\,\delta_4(P-P')\,z^2\,(2m)^{N_e'}\,|M|^2\,\frac{dp_1'}{(2\pi)^3}\ldots\delta(p_1'^2-m^2)\ldots.(7.39)$$

where N_e' is used to denote the number of final electrons. The invariant form of the cross section is found by noting that $\epsilon_1\epsilon_2|\mathbf{v}_1-\mathbf{v}_2|$ may be replaced by the invariant $|\{(p_1p_2)^2-p_1^2p_2^2\}|$. The cross section then appears in the form

$$d\sigma=\frac{(2\pi)^4\,\delta_4(P-P')}{|(p_1p_2)^2-p_1^2p_2^2|}\,\frac{(2m)^{(N_e+N_e')}}{4}\,|M|^2\frac{dp_1'}{(2\pi)^3}\ldots\frac{dp_{N'}}{(2\pi)^3}\,\delta(p_1'^2-m^2).\quad\ldots\ldots(7.40)$$

The energy momentum of one of the final particles, say p_1', will be completely determined in terms of the others

$$p_1'=P-\sum_2^{N'}p_i'.\qquad\ldots\ldots(7.41)$$

Thus in equation (7.40) we can integrate over p_1' removing the δ-function in energy momentum, obtaining the final form

$$d\sigma=\frac{\pi\,(2m)^{(N_e+N_e')}}{2\,|(p_1p_2)^2-p_1^2p_2^2|}\,|M|^2\,\frac{dp_2'}{(2\pi)^3}\ldots\delta(p_1'^2-m^2)\,\delta(p_2'^2-m^2)\ldots\quad\ldots\ldots(7.42)$$

where p_1' is determined by equation (7.41). This invariant form of the cross section is often the easiest to use when only total cross sections are required as special methods can be used to integrate the δ-function products without great difficulty.

The differential cross section in its usual form is best obtained directly from equation (7.38). Ordinary momentum conservation means that $\mathbf{p}_1'$ say is completely determined. The remaining δ-function, expressing energy conservation, implies that the energy of one of the other particles, say particle 2, is uniquely determined. We may integrate over this using the expression $d\mathbf{p}_2'=p_2'^2dp_2'd\Omega_2'$ and thus obtain the familiar result for the cross section.

$$d\sigma = 2\pi \frac{z^2 z'^2 |M|^2}{|\mathbf{v}_1 - \mathbf{v}_2|} p_2'^2 \left(\frac{\partial P_0'}{\partial p_2'}\right)^{-1} \frac{d\Omega_2'}{(2\pi)^3} \frac{d\mathbf{p}_3'}{(2\pi)^3} \cdots . \qquad \ldots\ldots (7.43)$$

applying to a process in which particle 2 goes into the solid angle $d\Omega_2'$, the remaining particles into the momentum space elements indicated. P_0' is the final energy of the system expressed as a function of p_2', the angle of emission of particle 2, $\mathbf{p}_3'$, etc.

§ 8. Elementary Applications

8.1. *Compton Scattering*

The chief interest in the Feynman–Dyson solution for the S-matrix lies in its power in the consideration of general questions of convergence and renormalization, and in the deduction of higher order radiative effects. It is interesting first, however, to see the application of the method to some of the more elementary of problems of radiation theory. The first problem to be considered will be that of Compton scattering. In lowest order this results from the sum of the two graphs G_1, G_2 of figure 2. We shall use the form (7.43) for the cross section, deriving this for the case in which the photon is scattered into an element of solid angle $d\Omega$. The initial electron and photon momentum vectors are taken as $\mathbf{p}$, $\mathbf{k}$ respectively and the final vectors $\mathbf{p}'$, $\mathbf{k}'$. The cross section then appears as

$$d\sigma = \left(\frac{e^2}{4\pi}\right)^2 \frac{m^2}{\epsilon\epsilon'} \frac{\kappa'}{\kappa} \left(\frac{\partial P_0'}{\partial \kappa'}\right)^{-1} \frac{1}{v} |M|^2 d\Omega \qquad \ldots\ldots (8.1)$$

where v is the projected velocity of approach of the incident particles and

$$M = \bar{u}^{i'}(p')\{e'(\not{p}+\not{k}-im)^{-1}e+e(\not{p}-\not{k}'-im)^{-1}e'\}u^i(p). \qquad \ldots\ldots (8.2)$$

Here i and i' denote the electron spin states. If we specialize to the frame of reference in which the electron is initially at rest $\partial P_0'/\partial \kappa' = m\kappa/\epsilon'\kappa'$ so that equation (8.1) reduces to

$$d\sigma = \left(\frac{e^2}{4\pi}\right)^2 \frac{\kappa'^2}{\kappa^2} |M|^2 d\Omega. \qquad \ldots\ldots (8.3)$$

Generally we will want to average over the incident electron spin and sum over the final electron spin for which purpose we may use the ' spur ' technique developed in § 5.2. For such an average we find

$$|M|^2 = -\frac{1}{8m^2} \operatorname{Sp}[R(\not{p}+im)\bar{R}(\not{p}'+im)] \qquad \ldots\ldots (8.4)$$

where, for abbreviation, R is used for the expression in curly brackets in equation (8.2). $\bar{R}$ has been described previously and is, in this case, simply R in inverted order and with a minus sign. The spur evaluation can be carried out by standard methods. In general with $\not{a} = a_\mu \gamma_\mu$

$$\frac{1}{4} \operatorname{Sp}(a_1 a_2) = a_1 a_2$$

$$\frac{1}{4} \operatorname{Sp}(a_1 a_2 a_3 a_4) = (a_1 a_2)(a_3 a_4) - (a_1 a_3)(a_2 a_4) + (a_1 a_4)(a_2 a_3). \qquad \ldots\ldots (8.5)$$

These rules can be extended to the product of any even number of a matrices. The spur of the product of an odd number of a's vanishes.

It is also desirable to be able to average the cross section over the initial polarization direction and sum it over the final polarization direction of the photon. There are only two independent real transverse polarization directions, but Feynman has shown that the correct results are obtained by summing over all four polarization directions. This is an immediate consequence of the gauge invariance, combined with the supplementary condition. From these the theory should be invariant to a change $e \to e+\lambda k$. That this is so may be verified explicitly for the matrix element M of equation (8.2), using the obvious results of the form

$$k(p+k-im)^{-1} = 1 - (p+k-im)^{-1}(p-im).$$

Suppose then that M in its dependence on e_μ has the form $M_\mu e_\mu$. It follows from the above that $k_\mu M_\mu = 0$. For simplicity suppose the z-axis to be the direction of propagation of the photon, then from the above $M_z = M_0$ and the sum over the four polarization directions $M_x^2 + M_y^2 + M_z^2 - M_0^2$ yields the same result as the sum over the two transverse polarizations.

We shall not deal here with the detailed derivation of the Klein–Nishina formula which follows quite rapidly from the above considerations. The invariant form for the cross section may be seen to be

$$\sigma = \left(\frac{e^2}{4\pi}\right)^2 \frac{8m^2}{(pk)} \int dk' \delta(k'^2) \delta((p+k-k')^2 - m^2) \,|\, M \,|^2. \qquad \ldots\ldots (8.6)$$

8.2. *Møller Scattering*

The problem of electron–electron scattering will be considered in lowest (e^2) order to illustrate how exclusion principle effects are automatically allowed for by the formalism. There is now only one graph for the process (figure 3(a)). Two

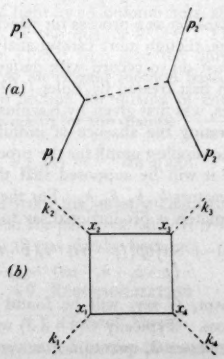


Figure 3. Graphs for (a) electron–electron scattering, (b) scattering of light by light.

electrons of momenta p_1, p_2 are scattered into momenta p_1', p_2' by the exchange of a virtual quantum. The initial state is described by a normalized state vector

$\langle(1/\sqrt{2})a^*(p_1)a^*(p_2)\,|\,0\rangle$ and there is a similar description of the final state. We require between these two states the matrix element of the second order component of S

$$S_2 = \tfrac{1}{2}e^2\int dx_1\int dx_2 T\{\bar{\psi}\gamma_\mu\psi A_\mu(x_1), \bar{\psi}\gamma_\lambda\psi A_\lambda(x_2)\}. \qquad \ldots\ldots(8.7)$$

In this case there are no virtual electron states. Each of the operators ψ must annihilate an incoming electron, each $\bar{\psi}$ must create an outgoing electron. There are four ways in which this can happen, but they divide into equivalent pairs of which typical members are

(*a*)	$\psi(x_1)$ annihilates p_1,	$\psi(x_2)$ annihilates p_2
	$\bar{\psi}(x_1)$ creates p_1',	$\bar{\psi}(x_2)$ creates p_2'
(*b*)	$\psi(x_1)$ annihilates p_1,	$\psi(x_2)$ annihilates p_2
	$\bar{\psi}(x_1)$ creates p_2',	$\bar{\psi}(x_2)$ creates p_1'.

On account of the anti-commutation properties of $a^*(p_1')\ a^*(p_2')$ there will be a difference in sign of the contribution to the S-matrix from the pairs (*a*) and (*b*).

Adding all the contributions and using the results of previous sections we arrive at the result

$$(p_1', p_2'\,|\,S_2\,|\,p_1, p_2) = (2\pi)^4\delta(p_1'+p_2'-p_1-p_2)\frac{ie^2}{2}\left(\frac{1}{2\pi}\right)^6\left(\frac{m^4}{\epsilon_1\epsilon_2\epsilon_1'\epsilon_2'}\right)^{1/2}$$
$$\left|\frac{[\bar{u}(p_1')\gamma_\mu u(p_1)][\bar{u}(p_2')\gamma_\mu u(p_2)]}{(p_1'-p_1)^2} - \frac{[\bar{u}(p_1')\gamma_\mu u(p_2)][\bar{u}(p_2')\gamma_\mu u(p_1)]}{(p_1'-p_2)^2}\right|. \qquad \ldots\ldots(8.8)$$

The expression for the differential cross section may be derived using the results of § 7.5, but will not be given here in detail.

8.3. *Scattering of Light by Light*

This is chosen for discussion as a process for which, at first sight the S-matrix element appears to diverge, though more careful analysis shows that the infinities cancel, as they indeed must do to accord with gauge invariance. The problem is an old one having been first treated by Euler (1936). A complete treatment, using covariant techniques, was first given by Karplus and Neuman (1950, 1951). The method of demonstrating the absence of infinities used here is essentially due to Ward (1950a). The simplest graph for the process is shown in figure 3 (*b*). For reasons of symmetry it will be supposed that the quanta are all incoming with energy momentum vectors k_1, k_2, k_3, k_4. For the graph shown in figure 3 (*b*) the contribution to the S-matrix is proportional say to $G(1\,2\,3\,4)$ where

$$G(1\,2\,3\,4) = e^4\mathrm{Sp}\int dp\{(p-im)^{-1}e_1(p+k_1-im)^{-1}e_2$$
$$\times(p+k_1+k_2-im)^{-1}e_3(p-k_4-im)^{-1}e_4\}. \quad \ldots.(8.9)$$

The total S-matrix element, G say, will be found by summing this over all permutations of the vertices. Typically $G(1\,4\,2\,3)$ will denote a graph in which quantum 4 is absorbed at vertex 2, quantum 3 at vertex 2 etc. Out of the total group of 24 permutations it is immediately evident that G is invariant to the cyclic group (1 2 3 4) and the sub-group (1 2) (3 4). There are thus essentially only three independent elements to be considered $G(1\,2\,3\,4)$, $G(1\,4\,2\,3)$, $G(1\,3\,4\,2)$.

We can now prove that G vanishes, either if any one of the k's vanishes, or if

one of the k's represents a longitudinal photon $e=\lambda k$. These are actually requirements for gauge invariance, and might be postulated. It is interesting, however independently to confirm their validity in this case. Suppose that $k_1=0$. Then by use of the identity

$$\operatorname*{Lim}_{k_1\to 0}(p-im)^{-1}(e_\mu\gamma_\mu)\,(p+k_1-im)^{-1}=-e_\mu\operatorname*{Lim}_{k_1\to 0}\frac{\partial}{\partial k_{1\mu}}(p+k_1-im)^{-1}$$

it may be verified that the sum of a typical pair, say

$$G(1234)+G(1432)\propto -e_{1\mu}\operatorname*{Lim}_{k_1\to 0}\frac{\partial}{\partial k_{1\mu}}\operatorname{Sp}\int dp\Big[\Big\{S_F(p+k_1)e_2\,S_F(p+k_2)e_3$$
$$\times S_F(p-k_4)e_4-S_F(p-k_1)e_4\,S_F(p+k_4)e_3\,S_F(p-k_2)e_2\Big\}\Big]$$

which is zero. Actually $G(1234)=G(1432)$ so both are zero separately. This condition is sufficient to show that the apparently divergent part of $G(1234)$ in fact vanishes. Equally if $e_1=\lambda k_1$, the identity

$$(p+k_1-im)^{-1}k_1(p-im)^{-1}=(p-im)^{-1}-(p+k_1-im)^{-1}$$

can be used to show that the sum of a pair $G(1432)+G(1324)=0$.

The convergence of the matrix element for the scattering of light by light and its gauge invariance have been demonstrated above. However, the actual calculation of the cross section is very elaborate and details will not be given here. The arguments above show that if G is expanded in powers of k the lowest order term occurring is of the type $k_1\,k_2\,k_3\,k_4$. The gauge invariance condition limits the forms in which this can appear in such a way that the corresponding interaction density for the electromagnetic field is of the form $a(F_{\mu\nu})^2+b(F_{\mu\nu}F_{\alpha\beta}\epsilon_{\mu\nu\alpha\beta})$ where $\epsilon_{\mu\nu\alpha\beta}$ is the isotropic totally antisymmetric tensor of fourth order.

The photon–photon scattering cross section is very small, of the order of 10^{-30} cm^2, and is unlikely to be observable in any circumstances. More accessible to observation is the scattering of light by a Coulomb field, also covered by figure 3 (*b*), though in this case for two of the quanta $k^2\neq 0$. This has been considered by Rohrlich and Gluckstern (1952) who give the differential cross section for forward scattering. They use an elegant method based on the optical theorem relating the absorption cross section for quanta of energy κ, to the (complex) forward scattering amplitude $a(\kappa,0)$ for the quanta

$$\sigma_{\text{abs}}(\kappa)=\frac{4\pi}{\kappa}\operatorname{Im}a(\kappa,0).$$

This result can be deduced from the unitary character of the S-matrix. The absorption cross section is that for pair creation. If it is known for all κ then by analytical continuation $\operatorname{Re}a(\kappa,0)$ can also be found.

§ 9. Renormalization

9.1. *General Survey*

For the simple processes treated in the last section the S-matrix elements have in each case been finite. However, it has been known for a long time that this ceases to hold when higher order radiative effects are calculated by direct use of the expansion (7.22). Typical of these divergent processes are the calculations of the self-energy of an electron, an early treatment of which according

to hole theory was given by Weisskopf (1934), or the polarization of the vacuum, early work on which is summarized by Weisskopf (1936). Great progress in an understanding of the difficulties has been made by a realization that all the infinities of electrodynamics are in fact associated with these two effects, and that they can be removed by appropriate infinite renormalizations of the electron mass and charge. It may be worth while to give a simple description of the application of these ideas to a typical case—the theory of the Lamb shift, essentially a displacement of the $2S_{1/2}$ level of the hydrogen atom caused by the interaction between the electron and the electromagnetic field (cf. Lamb and Retherford, 1950 and, for a survey of the experimental data, Salpeter 1953). This level shift simply represents the self-energy of the bound electron due to its ability, in interaction with the radiation field, to emit virtual photons, which may create pairs and so on. The calculated value of the self-energy even in the lowest order is infinite. However, it must be remembered that it has already been partially allowed for when, in calculation of the level, the observed mass m of the electron has been used in place of the bare mass m_0. What we should calculate is the energy difference between the self-energy of a bound electron and that of a free electron (of the same velocity). If this is done a finite value is found for the level shift. The first such calculations were carried out by Bethe (1947) following suggestions by Kramers (1948). Subsequent calculations of increasing refinement have since been carried out, and reference is made to these below. The Lamb shift includes also an effect due to polarization of the vacuum. In the Coulomb field of the proton virtual electron–positron pairs can be created and these will disturb the field acting on the bound electron. This disturbance, according to the calculations, multiplies the proton charge by an infinite constant, which must be absorbed in a renormalization of the coupling between the electron and proton. There are residual finite effects, first investigated by Uehling (1935), which lead to a further level shift.

The covariant forms of calculation described in the preceding sections are of great assistance in enabling the subtraction of infinities involved in renormalization to be carried out. The systematic use of such ideas is first found in papers by Tati and Tomonaga (1948) and Schwinger (1948). It was later shown by Dyson (1949a) that the renormalization could be carried through in all orders i.e. that with the definition of a renormalized mass m and charge e (involving of course infinite constants), to be identified with the empirically measured electron mass and charge, a finite value would be found for the S-matrix element in any order between any two states involving electrons and photons. Since then the technique of renormalization has been improved by various authors. Outstanding difficulties in the treatment by Dyson were cleared up by Ward (1951 a, b) and Salam (1951 a, b, 1952), and the methods were applied to the interaction of mesons, both with the electromagnetic and with the nucleon field, by various authors (particularly Salam 1951 a, b, 1952, Matthews 1950 a, b). In the treatment of Dyson, elimination of the infinities due to the self mass was accomplished by means of an infinite ' counter-term ' $\delta m\, \bar{\psi}\, \psi$ in the interaction Hamiltonian. Gupta (1951) extended the use of such infinite counter-terms to cover also charge renormalization. These ideas were developed by Takeda (1952) and Kallen (1952, 1953), the latter in particular showing the validity of the renormalization procedure without recourse to perturbation theory. A recent paper by Matthews and Salam (1954) gives a very clear account of the different methods of renormalization, and a concise statement of a procedure which eliminates the need for many

of the lengthy considerations of Dyson's earlier papers. Matthews and Salam (1951) have also written a survey article on the success of renormalization procedure in its application to meson theories.

It would be quite hopeless in a short review to try to give a complete description of all this work. The aim will rather be to give a general survey of the ideas of renormalization, and to show how they may be applied in practice to the calculation of higher order effects. Detailed effects dealt with in the next section are radiative corrections to scattering, the anomalous magnetic moment of the electron and the Lamb shift. The agreement with experiment shown by the calculations is good proof of the validity of the renormalization procedure, in spite of the manipulation of infinite quantities that is involved. Nevertheless one must hope that the subtraction of infinities will not be a permanent feature of radiation theory. Recent attempts, with theories explicitly using extended sources to avoid these infinities, have all run into insuperable physical difficulties. Another question raised by the S-matrix approach, proceeding as it does in a power series in the coupling constant, is whether, even after the individual terms have been made finite by renormalization, the series as a whole converges. Unfortunately this appears unlikely to be so. A simple scalar quantized field has been investigated by Thirring (1953) and Hurst (1952), and, for this case, the perturbation series expansion diverges for all values of the coupling constant. Physical arguments have also been given by Dyson (1952) making it plausible that the radius of convergence is zero also for perturbation theory in electrodynamics. Reference has already been made to these matters in the Introduction. A plausible attitude towards them is that the infinities will only be removed by going outside the domain of pure electrodynamics.

The general treatment of the renormalization problem for the S-matrix for interacting fields proceeds in three stages. Firstly, the number of types of infinity it contains (the primitive divergents of Dyson) is established. If this is finite the theory is renormalizable. A subtraction procedure has then to be defined to remove the infinities. Lastly, the dropping of the infinite terms has to be justified, and shown to correspond with a renormalization of the mass and coupling constant. We now consider these procedures for the case of electrodynamics; the results for meson theories will also be briefly referred to.

9.2. *Primitive Divergents*

Before we proceed further the interaction Hamiltonian will be modified to take account of mass renormalization. In the derivation of the Hamiltonian of (7.13) a Lagrangian for the bare electron field was used in which, tacitly, it was assumed that the mass m is that experimentally determined for the electron. In fact, in view of the self-energy, we should rather write $m=m_0+\delta m$, where m_0 is the mass of the bare particle and δm is the (in fact infinite) correction to the mass due to the interaction with the radiation field. Thus, where in the Hamiltonian density $m\bar{\psi}\psi$ has appeared it should be replaced by $m_0\bar{\psi}\psi=m\bar{\psi}\psi-\delta m\bar{\psi}\psi$. If the convention is to be retained that the bare electron field has the empirical mass m, then an extra term must be included in the interaction Hamiltonian, so that now

$$\mathcal{H}_i^1(x)=-j_\mu A_\mu-\delta m\bar{\psi}\psi. \qquad \ldots\ldots(9.1)$$

We now consider how divergences may arise in the evaluation of the S-matrix for any arbitrary graph for a process involving electrons and photons. Referring to § 7.4 the rules for the calculation of the S-matrix are that, apart from the external normalization factors, there must be included:

(i) at each vertex where a quantum, polarization μ, interacts with an electron, a factor $-ie\gamma_\mu(2\pi)^4\delta$(momentum) or, if the self-energy interaction acts, a factor $-\delta m(2\pi)^4\delta$(momentum), where the four-dimensional δ-function conserves energy and momentum at the vertex;

(ii) in order from the final to the initial state for each internal electron line a factor

$$-\left(\frac{1}{2\pi}\right)^4 \int_{C_F} S_F(p)dp \quad \text{with } S_F(p) = (p_\mu\gamma_\mu - im)^{-1};$$

(iii) for each internal photon line a factor

$$\left(\frac{1}{2\pi}\right)^4 \int_{C_F} D_F(k)dk \quad \text{with } D_F(k) = -i/k^2;$$

(iv) a factor $(-i)^n$ for the whole graph where n is the number of vertices, and a power of -1 determined by the considerations of § 7.4.

A typical graph is now considered which has say n vertices, F_e external and F_i internal fermion (electron) lines, and P_e external and P_i internal photon lines; also at n_s of the vertices the mass renormalization term acts. The divergence in the contribution from such a graph arises from the integration over the internal momenta p, k. This divergence may arise for three reasons. It may arise because the integrand does not vanish quickly enough for large p. It may arise from small values of p, or it may arise from coincidence of poles of the integrand. Divergences of the second type are well known as the infra-red catastrophe, which is really only a formal difficulty arising from the fact that our theory allows infinitely long electromagnetic waves, corresponding with zero energy quanta. It was shown by Bloch and Nordsieck (1937) how these difficulties may be avoided. Divergences of the third type are also not particularly interesting. It is the divergences of the first type, the old ' ultra-violet catastrophe ', which the renormalization procedure is designed to eliminate.

Any divergence of this type will be removed if a sufficient number of the internal momenta are held fixed, i.e. if in the corresponding graph a sufficient number of internal lines are cut, and thus made equivalent to a pair of external lines. A primitive divergent is defined as a divergent graph made convergent by the cutting of any internal line. For such graphs Dyson (1949 b) has shown that the integrals over the imaginary variables p_4 can be deformed into integrals over the real p_4 axis. The convergence of a graph can then be assessed by counting the powers of the different momentum components in the corresponding integral. Denoting by D the resultant power of all these momentum factors we see that the condition of convergence is $D<0$. From the rules for the formation of the S-matrix above it is simple to calculate the value of D. $\int S_F(p)\,dp$ gives a contribution of three powers from each internal electron line, and similarly two from each internal photon line. Each δ-function gives a contribution -4, and

$n-1$ of these are to be counted, as one is required for overall energy–momentum conservation. Thus

$$D=3F_i+2P_i-4(n-1). \qquad \ldots\ldots(9.2)$$

The construction of the graphs then shows that $2F_i+F_e=2n$; $2P_i+P_e=n-n_s$,

which together with (9.2) gives $D=4-\frac{3}{2}F_e-P_e-n_s$ $\ldots\ldots(9.3)$

and the convergence condition $\frac{3}{2}F_e+P_e+n_s>4$. $\ldots\ldots(9.4)$

Clearly there are only a very limited number of types of graph which fail to satisfy equation (9.4). The complete set is listed below.

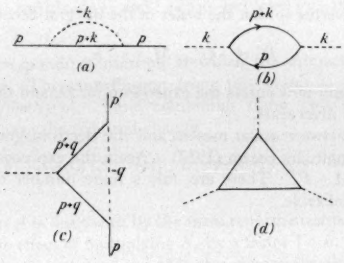


Figure 4. Primitive divergent graphs in electrodynamics.

(*a*) $F_e=2$, $P_e=0$. This is an electron self-energy graph of which the simplest type is shown in figure 4 (*a*). The possible degree of divergence is linear.

(*b*) $F_e=0$, $P_e=2$. This is a photon self-energy graph, figure 4 (*b*). The possible degree of divergence is quadratic, though, in fact, by gauge invariance, the quadratically divergent term must be identically zero.

(*c*) $F_e=2$, $P_e=1$. This is a vertex part, occurring for example in the radiative corrections to scattering or the Lamb shift. It is at most logarithmically divergent.

(*d*) $F_e=0$, $P_e=3$. This corresponds with the triangle diagram figure 4 (*d*). It can be proved in general that any closed electron polygon made up of an odd number of lines gives zero contribution to the *S*-matrix (Furry's theorem (Furry 1937)). The proof is relatively simple using the *S*-matrix methods, and follows from the fact that with any such graph there is associated a complementary one in which electrons and positrons are interchanged. This means simply that the direction of the electron line round the polygon is changed, and the complementary graph leads to a contribution equal to $(-1)^n$ times that of the original graph where n is the number of lines in the polygon.

(*e*) $F_e=0$, $P_e=4$. This is the case of scattering of light by light and could lead to a logarithmic divergence, but as has already been seen § 8.3, by gauge invariance it is in fact convergent.

There are, accordingly, just three primitive divergents in electrodynamics. Brief reference will be made here to the divergents of other theories. The meson–nucleon interaction with spin zero mesons is of the form $\bar{\psi}\Gamma\phi\psi$, where $\Gamma=\gamma_5$ for pseudoscalar mesons, $\Gamma=1$ for scalar mesons. This resembles electrodynamics in being a non-derivative coupling, and the dimensions D of the momentum integrals are given exactly as in electrodynamics by equation (9.3). The theories are thus renormalizable. However, the graphs (d) and (e) may now introduce divergences which have to be compensated by infinite counter terms, say $\lambda\phi^3$, $\mu\phi^4$, in the interaction Hamiltonian. Both these terms are necessary in the scalar theory, though, on account of Furry's theorem, only the latter is required for pseudoscalar mesons. With derivative type coupling, say $\bar{\psi}\gamma_5\gamma_\mu\psi\partial\phi/\partial x_\mu$, on the other hand, the interaction ceases to be renormalizable. There is now an added momentum at each vertex so that the order of the integral becomes

$$D=4-\tfrac{3}{2}F_e-M_e+n-n_s. \qquad \ldots\ldots(9.5)$$

The order of the graph now enters the expression for D, and there are an infinite number of primitive divergents.

The interaction between scalar mesons and the electromagnetic field has been considered, for example, by Salam (1952). Again the expression for D does not contain n: $D=4-M_e-P_e$. There are only a finite number of divergents, and the theory is renormalizable.

9.3. *Subtraction of Infinities*

Now that the possible divergences of the theory have been isolated the next stage is to devise a prescription for the removal of the consequent infinities. If we consider any arbitrary graph G then Dyson defines, associated with it, a skeleton graph G_0 formed by omitting all the self-energy and vertex parts from G. A graph which is its own skeleton is irreducible. Equally starting from any such irreducible graph, self-energy and vertex parts can be added in all possible ways. This leads to a set of graphs Γ say. If a prescription can be found for establishing how in terms of the contribution $M(G_0)$ of G_0 to the S-matrix the contribution $M(\Gamma)$ of Γ can also be found, then, in calculating the S-matrix, only irreducible graphs G_0 need be considered.

We begin very simply by considering the effect of the insertion of the elementary self-energy graph (figure 4 (a)) in an internal electron line of any graph. Using the rules already established it is easy to calculate that the effect of this insertion is to replace $S_F(p)$ in the original graph by

$$S_F(p)\frac{ie^2}{(2\pi)^4}\int\frac{dk}{k^2}\{\gamma_\mu S_F(p+k)\gamma_\mu\}S_F(p) = S_F(p)N(p)\,S_F(p), \text{ say.} \qquad \ldots\ldots(9.6)$$

Then, denoting $N(p)=e^2\int R(p,k)dk$, the integral for $N(p)$ is at most linearly divergent. The integrand $R(p,k)$ can be expanded in a Maclaurin series, expressed symbolically

$$R(p,k)=R(0,k)+p\frac{\partial}{\partial p}R(0,k)+R_c(p,k) \qquad \ldots\ldots(9.7)$$

where $\int R_c(p,k)dk$ converges. In this way, and using conditions of relativistic invariance, N can be uniquely expanded in the form

$$N=A+B(p\gamma-im)+(p\gamma-im)S_c(p) \qquad \ldots\ldots(9.8)$$

where A and B are infinite constants and $S_c(p)$ vanishes if $p\gamma=im$, i.e. if p satisfies the condition for a free electron. For a free particle N reduces to A, and by consideration of this case we can see that A simply gives, to order e^2, the change in mass of the electron due to its interaction with the radiation field. The only physical change that can occur for a free particle interacting with a radiation field is, indeed, a change of mass so that we would expect the modified propagator to be derived in this case from $S_F[p\gamma-i(m+\delta m)]$. Using the expansion

$$[p\gamma-i(m+\delta m)]^{-1}=(p\gamma-im)^{-1}+(p\gamma-im)^{-1}i\delta m(p\gamma-im)^{-1}+\ldots$$

we see that in fact $A=i\delta m$.

There is a corresponding insertion at which the interaction energy counter term $-\delta m\bar{\psi}\psi$ acts. The contribution from this graph is found by replacing $S_F(p)$ by $-iS_F(p)\delta mS_F(p)$. Thus combining these two insertions with the original graph we have the required change

$$S_F(p)\to S_F(p)+S_F(p)\{(A-i\delta m)+B(p\gamma-im)+(p\gamma-im)S_c(p)\}S_F(p). \qquad \ldots\ldots(9.9)$$

The infinity due to A is dealt with by the mass renormalization. The infinity due to B has simply the effect of multiplying S_F by a factor $1+B$. A consideration of the structure of the S-matrix shows that this renormalization, sometimes called wave function renormalization, is equivalent to a renormalization of the coupling constant, i.e. a charge renormalization. It could also be compensated by an infinite counter term in the interaction Hamiltonian of the form $\lambda\bar{\psi}(\gamma_\mu\partial/\partial x_\mu+m)\psi$. The simplest general proof of the practicability of the renormalization programme is in fact achieved in a form using such terms, together with others to compensate the remaining infinities described below (cf. Matthews and Salam 1954). For the present we are mainly concerned with the establishment of a technique for the separation of infinities.

Similar treatment may be given to the replacement of a typical vertex by the simple vertex part (figure 4 (*c*)). The vertex factor $-ie\gamma_\mu(2\pi)^4\delta(p'+k-p)$ is found to be modified by the replacement of γ_μ by $\Lambda_\mu(p,p')$ with

$$\Lambda_\mu(p,p')=-ie^2\left(\frac{1}{2\pi}\right)^4\int\frac{dq}{q^2}\{\gamma_\lambda S_F(p+q)\gamma_\mu S_F(p'+q)\gamma_\lambda\}. \qquad \ldots\ldots(9.10)$$

The integral is now at the most logarithmically divergent, and can be treated in the same way as (9.6) above. Using properties of relativistic covariance it can be seen that Λ_μ is of the form

$$\Lambda_\mu(p,p')=L\gamma_\mu+\Lambda_{\mu c}(p,p') \qquad \ldots\ldots(9.11)$$

where $\Lambda_{\mu c}$ is finite, and chosen so that it vanishes when $p=p'$ and p satisfies the free particle relation. For the simple case considered Λ is, of course, proportional to e^2. If one carries out the actual calculation of the second order contribution to Λ and S_F', equations (9.8) and (9.11), one finds that $L=-B$. Detailed calculations of this sort are given, for example, by Karplus and Kroll (1950). The term in L clearly

gives a second kind of charge renormalization, which, we see, in fact exactly compensates, to the order considered, the first kind (described as wave function renormalization). This is an example of a general identity proved by Ward (1950 b). Ultimately it can be shown to be due to the conservation of charge (i.e. to gauge invariance), as consideration shows to be the case in the present elementary example (Takeda 1952, Matthews and Salam 1954).

There is an immediate consequence of the identity when the radiative correction to electron scattering in lowest order is considered. Three of the graphs concerned are shown in figure 6. If mass renormalization is allowed for in the evaluation of the contribution from figure 6 (*b*) and (*c*) then the remaining divergent contributions from charge renormalization also cancel, and the sum of the contributions from the three graphs is finite. In cases such as this it is possible to introduce a cut-off factor in the integral over the virtual quantum energy in a covariant way in such a way that the cut-off parameter does not appear in the final result.

Finally, we deal with the photon self-energy part shown in figure 4 (*b*). If this is inserted in a photon line then the factor $D_F(k)=-i/k^2$ for the line is replaced by $D_F(k)\Pi(k)D_F(k)$ where

$$\Pi(k)=\frac{-e^2}{(2\pi)^4}\int dp \mathrm{Sp}\{\gamma_\lambda S_F(p)\gamma_\lambda S_F(p+k)\}. \qquad \ldots\ldots(9.12)$$

The integral is now quadratically divergent but again can be expanded to leave a finite remainder by the method of equation (9.7). The only relativistic invariant available for the expansion is k^2, so the expansion must be of the form

$$\Pi(k)=A'+Ck^2+k^2\Pi_c(k) \qquad \ldots\ldots(9.13)$$

where $\Pi_c(k)$ is a finite quantity vanishing when $k^2=0$.

The constant A', analogous to A of equation (9.8), is the photon self-energy which must be zero to preserve gauge invariance. If it were different from zero the required counter term in the interaction Hamiltonian would be of the form $\alpha\{A_\mu(x)\}^2$ which is contrary to gauge invariance. By suitable manipulation of the infinite integral defining A' Schwinger (1949 b) was able to make it plausible that formally A' should vanish. This type of argument is, however, open to criticism (Wentzel 1948) and it is better perhaps to use a regularization technique of the type discussed by Pauli and Villars (1949), which by covariant subtraction makes the photon self-energy directly zero. A technique of this sort is used also by Feynman (1949 b).

The second term in $\Pi(k)$ comes from a logarithmically divergent integral. It gives a renormalization of the electromagnetic field corresponding with an extra term of the type $\beta\{F_{\mu\nu}(x)\}^2$ in the interaction Lagrangian. Equivalently it can be regarded as a renormalization of charge. The remaining convergent term due to the polarization of the vacuum leads to observable effects and is more fully discussed below.

So far only the most simple additions to the lines of a graph have been considered. The ideas can, however, be generalized, and the recipe for obtaining the complete contribution to the S-matrix from the graphs Γ corresponding with an irreducible G_0 stated as follows. For every electron line we replace $S_F(p)$ by $S_F'(p)$ where $S_F'(p)=S_F(p)+S_F(p)\Sigma(p)S_F(p)$, $\Sigma(p)$ denotes a sum of contributions

such as N above, but now over all possible self-energy parts. Similarly the photon propagator is replaced by $D_F'(p)$, and the matrix γ_μ occurring at a vertex by Γ_μ where

$$D_F'(p)=D_F(p)+D_F(p)\,\Pi(p)D_F(p)$$

$$\Gamma_\mu(p,p')=\gamma_\mu+\Lambda_\mu(p,p'). \qquad \ldots\ldots(9.14)$$

If we restrict the self-energy and vertex parts added to be themselves irreducible and ' proper ' (i.e. not divisible into two pieces joined by a single line) then the degree of divergence of the corresponding $\Sigma^*(p)$, $\Pi^*(p)$, $\Lambda_\mu^*(p,p')$ is exactly as for the simple parts already considered, and, after certain infinite terms corresponding with charge and mass renormalizations have been removed, the remainders are finite. If, with attention still confined to proper self-energy parts the requirement of reducibility is removed then the complete modified propagators can be defined by equations of the form

$$S_F'(p)=S_F(p)+S_F(p)\Sigma^*(p)S_F'(p). \qquad \ldots\ldots(9.15)$$

A prescription has then, however, to be given for the removal of divergences from the contribution to $\Sigma^*(p)$ say from electron self-energy parts containing vertex parts or other self-energy parts within their structure. This has been given by Dyson, amplified by Salam (1951 a, b). An alternative treatment is given by Ward (1951 a, b). Little trouble arises from a reducible vertex part V_R with various inserted self-energy and vertex parts. The integral giving the contribution of V_R diverges, not only as a whole, but also when integrated over the variables corresponding to the insertions. The divergences of V_R are removed in succession, according to the rules already laid down, beginning with the insertions and finishing with the divergences arising from V as a whole.

Reducible self-energy parts are more complicated on account of the overlapping divergences that then arise. For example, in figure 5 are shown graphs

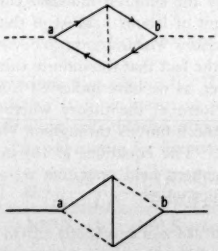


Figure 5. Overlapping divergent graphs.

made up by adding a vertex part either to an electron or to a photon self-energy part. This vertex must be regarded as added at either vertex a or vertex b but not at both to avoid duplication of results. The complete prescription for dealing with such divergences is given by Salam (1951 a, b).

Using these subtraction techniques, Dyson was able to show that the finite contributions S_{F1}', D_{F1}', $\Gamma_{\mu 1}$ left after the subtraction of infinities were related to the complete quantities by multiplicative relations

$$\Gamma_\mu = Z_1^{-1}\,\Gamma_{\mu 1}(e_1)$$
$$S_F' = Z_2 S_{F1}'(e_1)$$
$$D_F' = Z_3 D_{F1}'(e_1)$$
$$\text{with} \qquad e_1 = Z_1^{-1} Z_2 Z_3^{1/2} e. \qquad \ldots\ldots(9.16)$$

On account of Ward's identity in fact it appears that $Z_1 = Z_2$; e_1 is the renormalized charge, as compared with the unobservable bare charge e. The application of the renormalization method to the evaluation of the S-matrix elements, including radiative corrections, for some convergent process can now be described. First the irreducible graphs for the process are drawn. A typical one of these of order n will contribute a convergent integral with integrand of the form $e^n f(\gamma_\mu, S_F, D_F)$. The radiative corrections to this graph are introduced by replacing γ_μ by Γ_μ, S_F by S_F', D_F by D_F', and the integral now diverges. However,

$$e^n f(\Gamma_\mu, S_F', D_F') = (e Z_1^{-1} Z_2 Z_3^{1/2})^n f(\Gamma_{\mu 1}, S_{F1}', D_{F1}')$$
$$= e_1^n f(\Gamma_{\mu 1}, S_{F1}', D_{F1}') \qquad \ldots\ldots(9.17)$$

so that the divergent integrand is just an infinite multiple of an integrand leading to a convergent result. This leads to the identification of e_1 with the empirically measured charge, and a finite result is found for the graph considered with all its added radiative corrections. Thus finally we see the triumph of renormalization theory which isolates all the infinities of electrodynamics in two infinite constants the values of the mass and charge of the electron.

The ideas of the renormalization procedure probably appear most clearly in the step-by-step elimination of infinities from the different graphs of the Dyson method described above. Also the results are found in a method well adapted for application to the calculation of low order radiative corrections to real processes. In the next section, where a few such problems are considered, it will become very evident that in practice only the simplest radiative corrections can be considered without a prohibitive amount of labour. Most of the involved discussion of the method is required for processes which are quite beyond the range of calculation, but it is required to justify the fact that the infinite subtractions do merely involve a renormalization. However, as we have indicated from time to time this general proof is better given in a form of the theory where infinite counter terms are introduced to compensate the infinities throughout (Matthews and Salam 1954, Kallen 1953, Gupta 1951). The equations (9.16) indicate how this should be done. Renormalized Heisenberg field operators ψ_1, A_1 are introduced into the Lagrangian of electrodynamics where

$$\psi = Z_2^{1/2}\psi_1, \quad A = Z_3^{1/2} A_1$$
$$\text{with } e = Z_1 Z_2^{-1} Z_3^{-1/2} e_1, \quad \text{and } m = m_0 + \delta m. \qquad \ldots\ldots(9.18)$$

Expressing $\mathcal{L}$ in terms of these variables one finds

$$\mathcal{L} = \mathcal{L}_{10} + \mathcal{L}_1^{\,1} \qquad \ldots\ldots(9.19)$$

where $\mathcal{L}_{10}$ has the form of the bare field Lagrangian, but in terms of the renormalized operators, and

$$\mathcal{L}_1^{\,1} = -ie_1[1 - L(e_1)] A_{1\mu} \bar{\psi}_1 \gamma_\mu \psi - B(e_1)\bar{\psi}_1 \left(\gamma_\mu \frac{\partial}{\partial x_\mu} + m\right)\psi_1$$
$$+ Z_2 \delta m \bar{\psi}\psi - \tfrac{1}{4} C(e_1)\,(F_{1\mu\nu})^2 \qquad \ldots\ldots(9.20)$$

where we have defined

$$Z_1=1-L(e_1)\,,\quad Z_2=1+B(e_1)\,,\quad Z_3=1+C(e_1). \qquad \ldots\ldots(9.21)$$

The new renormalized interaction representation is introduced from $\mathcal{L}_{10}$, and the corresponding interaction representation field variables now create particles of the true mass and charge. The counter terms in $\mathcal{L}_1^{\mathrm{I}}$, with appropriate definition of the constants, will compensate all the infinities arising in the theory.

§ 10. Calculation of Radiative Effects

10.1. *Introduction*

In this section a brief summary will be given of some of the effects due to higher order radiative corrections which have been the subject of calculation. No new difficulties of principle are introduced by such calculations. The Dyson procedure (or any of its equivalent variants) can in theory be applied to find any modified propagation function required, say $S_{\mathrm{F}}'(p)$, to any order in e^2. However, the practical difficulties of such calculations are enormous and, to take an example, in the case of Compton scattering there are still unresolved discrepancies in the published values for the (lowest order) correction to the Klein–Nishina formula. The methods of calculation for higher order effects proposed by Feynman (1949 b) are in many ways the simplest to apply in practice, and the description below is generally given in terms of his technique.

10.2. *Electron Self-Energy*

We begin with a consideration of the electron self-energy. From the considerations of § 9.3 it appears that to order e^2 the electron self-mass, due to interaction with the radiation field, is given by

$$\delta m=\frac{e^2}{(2\pi)^4}\int\frac{dk}{k^2}\{\gamma_\mu S_{\mathrm{F}}(p+k)\gamma_\mu\} \qquad \ldots\ldots(10.1)$$

evaluated for a state u for which $(\gamma p-im)u=0$. This integral diverges, and Feynman proposes to regularize it in a covariant way by replacing $1/k^2$ by $\{1/k^2-1/(k^2+\lambda^2)\}$. The self-mass will depend upon the cut-off energy λ becoming infinite as $\lambda\to\infty$. The expression for δm may be written

$$\delta m=\frac{e^2}{(2\pi)^4}\int_0^{\lambda^2}dL\int\frac{dk}{(k^2+L)^2}\,\frac{\gamma_\mu\{\gamma(p+k)+im\}\gamma_\mu}{k^2+2(pk)} \qquad \ldots\ldots(10.2)$$

where in the denominator of the integrand use has been made of the fact that $p^2=m^2$. Feynman then proposes a technique for the reduction of such integrals to a standard form, by use of the result

$$ab=\int_0^1\frac{dx}{\{ax+b(1-x)\}^2} \qquad \ldots\ldots(10.3)$$

and similar expressions derived by differentiating with respect to a or b. In this way the integral for δm is expressed in the form

$$\frac{e^2}{(2\pi)^4}\int_0^{\lambda^2}dL\int_0^1 2x\,dx\int dk\,\frac{\gamma_\mu[\gamma(p+k)+im]\gamma_\mu}{\{k^2+2(1-x)(pk)+xL\}^3}. \qquad \ldots\ldots(10.4)$$

The denominator of expression (10.4) can be brought to normal form, and finally the use of the integral

$$\int \frac{dk}{(k^2+L)^3} = \frac{i\pi^2}{2L}$$

enables the whole expression to be evaluated leading to the result

$$\frac{\delta m}{m} = \frac{e^2}{4\pi}\left(3 \log \frac{\lambda}{m} + \frac{3}{4}\right) \qquad \ldots\ldots(10.5)$$

for the mass correction to order e^2. This value of δm is the one, according to the present procedure, which must be used in the subtraction term $\delta m \bar{\psi}\psi$.

It is interesting to contrast this result with that obtained by following through the procedure of Dyson (1949 b) according to which the integral (10.1) is to be expanded in the form $A+B(\gamma p - im)+$ terms of order $(\gamma p - im)^2$ where A and B are both infinite, and to be compensated respectively by charge and mass renormalization. A calculation of this sort is given by Karplus and Kroll (1950). δm now appears formally as a divergent integral, as does the charge renormalization term. The finite correction to S_F in second order given by these authors is very lengthy, and will not be quoted here. This finite correction would, however, equally be arrived at in the formal method of Feynman as a term independent of 'cut-off' λ. It vanishes for a free particle, and hence is only required for the evaluation of graphs with electron self-energy parts in internal lines, none of which will here be considered in detail.

10.3. *Anomalous Electron Magnetic Moment*

We now consider the effect of radiative corrections on the interaction between an electron and an external electromagnetic field. This leads to three interesting and important problems, in which finite effects are found due to radiative corrections, the first of which to be considered will be the anomalous magnetic moment of the electron. In this case the interaction is with a weak magnetic field. There will, therefore, be no important error in regarding the electron as free both in its initial and intermediate state, and the methods developed above can be immediately applied. The effect was first calculated by Schwinger (1949 b). An alternative derivation is due to Geheniau and Villars (1950).

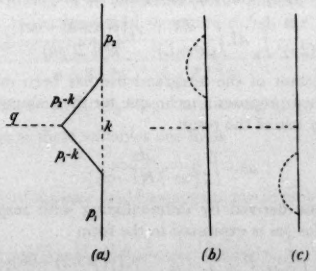


Figure 6. Graphs for radiative correction to scattering.

We consider a graph of the type shown in figure 6 (*a*) where an electron of momentum p_1 emits a virtual quantum k, interacts with an external field, receiving momentum q, and after reabsorbing the virtual quantum has momentum p_2. The external field will be denoted by

$$A_\mu^e(x) = \left(\frac{1}{2\pi}\right)^4 \int A_\mu^e(q) e^{iqx}\, dq. \qquad \ldots\ldots(10.6)$$

The graph 6 (*a*) is a particular case of the vertex part considered in § 9.3, and its effect in general is described by the addition to the γ_μ, occurring at the vertex of a single interaction with the electromagnetic field, of a matrix $\Lambda_\mu(p_1, p_2)$ which contains an infinite multiple of γ_μ, to be absorbed by charge renormalization, in addition to a finite part here required to order e^2. This matrix is again evaluated by Karplus and Kroll (1950). We require its form only when $p_1^2 = p_2^2 = m^2$ and sketch its derivation by the regularization procedure of Feynman.

The first order matrix element for interaction with the external field A_μ^e, apart from renormalization factors has the form $-e\gamma_\mu A_\mu^e(p_2 - p_1)$. The radiative correction consists in the replacement of γ_μ by

$$\Lambda_\mu^{(2)} = -\frac{i\alpha}{4\pi^3}\int \frac{dk}{k^2}\{\gamma_\nu S_F(p_2-k)\gamma_\mu S_F(p_1-k)\gamma_\nu\} \qquad \ldots\ldots(10.7)$$

where $\alpha = e^2/4\pi$ denotes the fine structure constant. The integral now diverges at both limits. We remove the infra-red divergence by introducing a small mass λ_0 for the photons, and the high energy divergence by the use of the same convergence factor as in equation (10.2). This leads to an expression

$$\Lambda_\mu^{(2)} = -\frac{i\alpha}{4\pi^3}\int_0^{\lambda^2} dL \int dk \frac{\gamma_\nu[\gamma(p_2-k)+im]\gamma_\mu[\gamma(p_1-k)+im]\gamma_\nu}{(k^2+\lambda_0^2)\{(p_2-k)^2+m^2\}\{(p_1-k)^2+m^2\}} \qquad \ldots\ldots(10.8)$$

which can be evaluated by the methods of the previous section. The result is of the form

$$\Lambda_\mu^{(2)} = -R\gamma_\mu + \Lambda_{\mu c}^{(2)}(q),\ q = p_2 - p_1, \qquad \ldots\ldots(10.9)$$

where R depends on the cut-off, becoming infinite as $\lambda \to \infty$,

$$R = \log\frac{\lambda}{m} + \frac{9}{4} - 2\log\frac{m}{\lambda_0} \qquad \ldots\ldots(10.10)$$

and $\Lambda_{\mu c}^{(2)}(q)$ vanishes when $q^2 = 0$, and is given by

$$\Lambda_{\mu c}^{(2)}(q) = \frac{\alpha}{2\pi}\{-A\gamma_\mu + B(q\gamma_\mu - \gamma_\mu q)\} \qquad \ldots\ldots(10.11)$$

where

$$A = 2(\log\frac{m}{\lambda_0} - 1)(1 - 2\zeta\cot 2\zeta) + \zeta\tan\zeta + 4\cot 2\zeta\int_0^\zeta x\tan x\, dx$$

$$B = \frac{i}{4m}\zeta\operatorname{cosec} 2\zeta, \quad \text{where } \sin^2\zeta = -\frac{1}{4}q^2. \qquad \ldots\ldots(10.12)$$

In deriving the above, terms in λ_0/m, m/λ have been omitted, and use made of $p_1^2 = p_2^2 = m^2$. A treatment by the method prescribed by Dyson without this last restriction is again given by Karplus and Kroll (1950).

The term in R in equation (10.9) is removed by charge renormalization. On

account of Ward's identity (§ 9.3) it is in fact compensated by similar terms introduced when in the single scattering, figure 6 (*b*), (*c*), self-energy parts are introduced in the electron lines, so that in this case no reference is in fact needed to charge renormalization (cf. Lewis 1948).

Turning now to the finite terms we see that the anomalous magnetic moment must arise from the terms in (10.11) proportional to B. In the limit of small q^2 these yield a contribution to the S-matrix of the form, again aside from external spinors and normalization factors

$$-ie\,A_\mu^{\,e}(q)\,\frac{\alpha}{8\pi m}(q\gamma_\mu-\gamma_\mu q). \qquad \ldots\ldots(10.13)$$

This must be compared with the contribution which would arise from the action of an anomalous magnetic moment, say $\mu e/2m$. This latter would have an interaction energy density $-\mu(e/2m)\frac{1}{2}\sigma_{\mu\nu}F_{\mu\nu}^{\,e}$ where $\gamma_\mu\gamma_\nu-\gamma_\nu\gamma_\mu=2i\sigma_{\mu\nu}$. The corresponding first order S-matrix element is

$$-\frac{\mu e}{2m}\,\sigma_{\nu\mu}\,q_\nu\,A_\mu^{\,e} \qquad \ldots\ldots(10.14)$$

and comparison of this with (10.13) shows that $\mu=\alpha/2\pi$. Karplus and Kroll (1950) were able to take this calculation to the fourth order, finding an extra contribution, in units $e/2m$ amounting to $-2{\cdot}973\alpha^2/\pi^2$. Recent precise experiments (Koenig, Prodell and Kusch 1952) appear to confirm the need for this fourth-order correction.

10.4. *Radiative Corrections to Electron Scattering*

We consider scattering by a Coulomb field in Born approximation. If the scattering centre has charge $-Ze$ then to first order in Ze the S-matrix element for the scattering can be immediately calculated as

$$S^{(1)}=\frac{iZe^2}{4\pi|\mathbf{p}_2-\mathbf{p}_1|^2}\frac{1}{\pi}\bar{u}^{(}(p_2)\gamma_4 u^{i'}(p_1)\frac{m}{(\epsilon_1\epsilon_2)^{1/2}}\,\delta(\epsilon_1-\epsilon_2), \qquad \ldots\ldots(10.15)$$

where $(\mathbf{p}_1,\ \epsilon_1)$, $(\mathbf{p}_2,\ \epsilon_2)$ denote the energy and momenta of the initial and final electron states. From equation (10.15) the expression for the cross section can be derived for scattering of the electron into a solid angle $d\Omega$

$$d\sigma=4\,Z\left(\frac{e^2}{4\pi}\right)^2\frac{1}{|\mathbf{p}_2-\mathbf{p}_1|^4}\frac{m^2p}{\epsilon}\,|\,\bar{u}^{(}(p_2)\gamma_4 u^{i'}(p_1)\,|^2 d\Omega \qquad \ldots\ldots(10.16)$$

where, for the moment, $p=|\mathbf{p}_1|=|\mathbf{p}_2|$, $\epsilon=\epsilon_1=\epsilon_2$. Summing and averaging over spins gives the usual Born approximation for relativistic electrons

$$d\sigma=\frac{Z^2(e^2/4\pi)^2}{4p^2v^2\sin^4\frac{1}{2}\theta}(1-v^2\sin^2\tfrac{1}{2}\theta). \qquad \ldots\ldots(10.17)$$

The effect of radiative corrections, from the work of the previous section, is simply to replace γ_4 in equation (10.15) or (10.16) by $\gamma_4+\Lambda_{4c}(q)$, which is given to the second order in equation (10.11) above. In the general expression we must put $\mathbf{q}=\mathbf{p}_2-\mathbf{p}_1$, $q_0=0$. The evaluation of the cross section to order e^6 is then quite direct but rather tedious. Details will be found in papers by Schwinger (1949 b), and Fukuda, Miyamota and Tomonaga (1949) who were the first to calculate the effect. One feature of the calculations, however, deserves comment. A fictitious

photon mass λ_0 was introduced in equation (10.8) to avoid the infra-red catastrophe, and λ_0 will still appear in the expression for the radiative correction to the cross section. The difficulty, being one associated with very long electromagnetic waves, has clearly nothing to do with relativistic quantum mechanics. It was first analysed by Bloch and Nordsieck (1937). Later treatments by Pauli and Fierz (1938), and Dancoff (1939) made it clear that the divergent negative radiative correction to the elastic scattering cross section from virtual photons of energy less than some value K is exactly compensated by the cross section for scattering with emission of radiation of energy less than K. In the form of the theory here presented, if the cross sections for elastic scattering, and for inelastic scattering, accompanied by the emission of a photon of energy up to K, are added then λ_0 disappears from the total cross section which now depends rather upon K. Experimentally it is only the sum of these cross sections that can be measured. The divergence of the individual cross sections is unimportant, being merely associated with the fact that an infinite number of low-energy photons will be emitted in any collision of charged particles.

Calculations of radiative corrections to scattering have been extended by Redhead (1953) to the case of electron–electron and electron–positron scattering. To obtain the S-matrix up to order e^4, graphs in which two photons are exchanged between the electrons have to be included, together with the radiative corrections. This is equivalent to the inclusion of two scatterings by the Coulomb field in the previous example. Also the effect of polarization of the vacuum by the photon exchanged between the electrons must be included. This was ignored in the consideration of scattering by a Coulomb field, but it is described in more detail below in connection with the Lamb shift. Redhead presents his results very conveniently in a way which separates out the contributions to the radiative corrections from the different effects. Numerically his results show that for an incident energy of say 10 MeV the correction to the cross section leads to its diminution by several per cent. Experiments have not yet been carried out of sufficient precision to verify this effect.

10.5. *The Lamb Shift*

The Lamb shift has so far provided the best test of radiative correction phenomena on account of the precise comparison that in this case can be made with experiment. What is measured is the energy difference between the $2S_{1/2}$ and $2P_{1/2}$ levels of hydrogen. Recent experimental results (Triebwasser, Dayhoff and Lamb 1953) give this with an accuracy to 1 in 10^4 as $1057{\cdot}77 \pm 0{\cdot}10$ Mc/s. The levels are coincident in the one-electron Dirac theory, so that the complete level splitting is to be ascribed to the relative shift produced by radiative effects. The phenomenon has already been described in a general way in § 9.1. The major part of the shift comes from a non-relativistic contribution, applying only to the 2S state, first calculated by Bethe (1947). This contribution is of order $\alpha(Z\alpha)^4 \log Z\alpha$ as compared with an, in fact considerably smaller, relativistic contribution of order $\alpha(Z\alpha)^4$. Early relativistic treatments were given by French and Weisskopf (1949) and Feynman (1949 b). Recent treatments by Karplus, Klein and Schwinger (1952) and Baranger, Bethe and Feynman (1953) have investigated more carefully the fitting together of the relativistic and non-relativistic regions, obtaining also corrections of order $Z\alpha$ relative to the main contributions. Agreement with

experiment is now within 0·6 Mc/s, and this remaining discrepancy may be accounted for by still higher order corrections. Various other sources of error are discussed by Salpeter (1953).

Contributions to the Lamb shift arise from two sorts of graph. Firstly there is the electron self-energy type of graph figure 6 (*a*), already considered in connection with radiative corrections to scattering and the anomalous magnetic moment. Secondly there is the polarization of the vacuum type of graph illustrated in figure 7, which though it contributes to the scattering problem has so far been

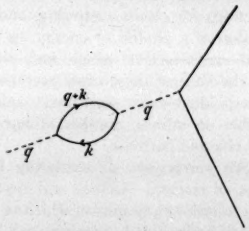


Figure 7. Graph for polarization of vacuum.

ignored. We shall begin by considering the first type of graph. In figure 6 (*a*) the lines were assumed to denote electrons propagating freely, and the effect of the Coulomb field was only included in the first Born approximation, as one interaction with the external field for the intermediate state electron. It might be thought to correct this approximation by the replacement of the initial and final plane wave electron states with the appropriate bound state wave function. In the calculation of the vertex part correction, admittedly, use was made of the fact that $p_1^2=p_2^2=m^2$, but the deviation from unity of p^2/m^2 for a bound state (with nuclear charge Z) will only be of order $(Z\alpha)^2$ which is small. This corrected approximation, however, will only have a validity when it is permissible to treat the intermediate electron wave function as determined by the first Born approximation. There will be a region in which this is so—the relativistic region referred to above—when the energy κ of the virtual photon is above some limit, say $Z\alpha m$. Below this limit it will be possible to give a simple non-relativistic treatment of the type originally given by Bethe. The intermediate electron states are then represented either by bound atomic states or states of the continuum specified by their non-relativistic wave functions, and the calculation is most easily carried out directly by means of non-covariant perturbation theory.

The above approximations are of a rather hybrid nature, which is due to the fact that we have been using an interaction representation based on free particle states to treat electrons moving in the field of a nucleus. The formal difficulties will be removed if we introduce a 'bound state interaction representation' in which the electron field operators satisfy an equation

$$\gamma_\mu\left(\frac{\partial}{\partial x_\mu}-ie\,A_\mu^e(x)\right)\psi+m\psi=0 \qquad \ldots\ldots(10.18)$$

where the external field is denoted by $A_\mu^e(x)$. The operators ψ can then be

expanded in terms of annihilation operators and the c-number solutions of (10.18), say

$$\psi=\Sigma A_r\phi_r(x). \qquad \ldots\ldots(10.19)$$

Such bound interaction representations were first considered by Furry (1951). Their renormalization is discussed briefly by Matthews and Salam (1954), who show that it is validly carried out with the use of the 'free' interaction representation counter terms. The construction of the S-matrix in the bound representation can be carried through just as in § 7.4. However, the electron propagation function $S_F^e(x', x)$ defined as previously by

$$\tfrac{1}{2}S_F^e(x', x) = \langle T(\psi(x'), \bar{\psi}(x))\rangle_0 \qquad \ldots\ldots(10.20)$$

is no longer a function of the interval $x'-x$, since the external field is localized in space-time. If the c-number solutions of (10.18) are denoted by $\phi_n(x)$ and normalized so that $\int\phi_m^*(x)\phi_n(x)dx=\delta_{mn}$, then it is easily verified that

$$\tfrac{1}{2}S_{F\,\alpha\beta}^e(x', x)=\sum_{E_n>0}\{\phi_n(x')\}_\alpha\{\bar{\phi}_n(x)\}_\beta,\ t'-t>0 \quad \ldots.(10.21)$$

$$=-\sum_{E_n>0}\{\phi_n(x')\}_\alpha\{\bar{\phi}_n(x)\}_\beta,\ t'-t<0.$$

The shift in the level of some bound state due to the radiative interaction may then be found from the expectation value of the S-matrix for this state. It will be sufficient to work only to the second order in S i.e. to consider only

$$S^{(2)}=-\frac{e^2}{4}\int dx'\int dx\ T\{\bar{\psi}\gamma_\mu\psi(x'),\ \bar{\psi}\gamma_\mu\psi(x)\}D_F(x'-x). \quad \ldots\ldots(10.22)$$

If the bound one-electron state is denoted a then we require $(a\,|\,S^{(2)}\,|\,a)$. Corresponding to an interaction energy $\mathcal{H}_s(x)$ there is an S-matrix contribution, in lowest order, $-i\int dx\mathcal{H}_s(x)$. To the order required the change in energy is thus given by

$$\int H_s(t)dt=\mathrm{Re}\{i(a\,|\,S_2\,|\,a)\} \qquad \ldots\ldots(10.23)$$

where account has still to be taken of mass and charge renormalization. Strictly the expression (10.23) with $S^{(2)}$ defined in (10.22) includes both the electron self-energy and the vacuum polarization part of the Lamb shift. The latter contribution arises when operators at one world point are allowed to create and annihilate the same electron. This process gave zero contribution in the absence of an external field but has now to be considered. For the moment we ignore it. The level shift ΔE_a then appears as

$$\Delta E_a=\mathrm{Re}\left[-\frac{ie^2}{4}\int d\mathbf{x}'\int d\mathbf{x}\int d(t'-t)\bar{\phi}_a(x')\gamma_\mu S_F^e(x', x)\gamma_\mu\phi_a(x)D_F(x'-x)\right]$$

$$\ldots\ldots(10.24)$$

together with the counter term $-\delta m\int\bar{\phi}_a(x)\phi_a(x)d\mathbf{x}$. It may be noted that if the expression (10.21) is introduced for S_F^e the older form of the result for ΔE_a (cf. French and Weisskopf (1949)) will be reproduced.

The momentum representation will now be introduced with the following definitions

$$\phi_a(x)=\phi_a(\mathbf{x})\exp(-iE_at),\quad \phi_a(\mathbf{p})=\left(\frac{1}{2\pi}\right)^{3/2}\int\phi_a(\mathbf{x})\exp(-i\mathbf{p}.\mathbf{x})d\mathbf{x}$$

$$A_\mu^e(x)=\left(\frac{1}{2\pi}\right)^3\int A_\mu^e(\mathbf{q})\delta(q_0)\exp(iqx)dq \qquad \ldots\ldots(10.25)$$

so that it is assumed the external field is constant in time.

$$\tfrac{1}{2}S_F^e(x',x)=\left(\frac{1}{2\pi}\right)^4\int S_F^e(p',p)\delta(p_0'-p_0)\exp\{i(p'x'-px)\}dp\ dp'\ldots\ldots(10.26)$$

where use is made of the fact that S_F^e depends on t and t' only in the combination $t'-t$. If these definitions are introduced into equation (10.24) one obtains, with the understanding that the real part is to be taken

$$\Delta E_a=-\frac{e^2}{(2\pi)^4}\int\frac{dk}{k^2}\int d\mathbf{p}'\int d\mathbf{p}\ \phi_0(\mathbf{p}')\gamma_\mu S_F^e(\mathbf{p}'-\mathbf{k},\mathbf{p}-\mathbf{k};E_a-k_0)\gamma_\mu\phi_0(\mathbf{p}).$$

$\ldots\ldots(10.27)$

We must now investigate the form of S_F^e. This may be found from a consideration of the equation

$$\left(\gamma_\mu\frac{\partial}{\partial x_\mu'}+m\right)S_F^e(x',x)=ie\gamma_\mu A_\mu(x')S_F^e(x',x)+2i\delta(x'-x),\ \ldots\ldots(10.28)$$

the comparable equation involving x, and the causal properties of S_F. The Fourier transform of (10.28) yields

$$(\gamma_\mu p_\mu'-im)S_F^e(p',p)=\frac{e\gamma_\mu}{(2\pi)^3}\int A_\mu^e(\mathbf{p}'-\mathbf{p}'')S_F^e(p'',p)d\mathbf{p}''+\delta(\mathbf{p}'-\mathbf{p}).$$

$\ldots\ldots(10.29)$

This may be written symbolically as a matrix equation

$$\left(\gamma p-im-\frac{e\gamma_\mu A_\mu^e}{(2\pi)^3}\right)S_F^e=1$$

which makes it very direct in terms of the expansion

$$C=(A+B)^{-1}=A^{-1}+A^{-1}BA^{-1}+A^{-1}BA^{-1}BA^{-1}+\ldots$$

to derive

$$S_F^e(p',p)=(\gamma p-im)^{-1}\delta(\mathbf{p}'-\mathbf{p})+(\gamma p'-im)^{-1}\frac{e\gamma_\mu}{(2\pi)^3}A_\mu^e(\mathbf{p}'-\mathbf{p})(\gamma p-im)^{-1}$$

$$+(\gamma p'-im)^{-1}\frac{e\gamma_\mu}{(2\pi)^3}A_\mu^e(\mathbf{p}'-\mathbf{p}'')S_F^e(p'',p''')\frac{e\gamma_\mu}{(2\pi)^3}A_\mu^e(\mathbf{p}'''-\mathbf{p})(\gamma p-im)^{-1}$$

$\ldots\ldots(10.30)$

where in the last term an integral is assumed over the repeated variables $\mathbf{p}''$, $\mathbf{p}'''$.

Inserting equation (10.30) in the expression for ΔE_a we find, in view of (10.1), that the first term exactly cancels the self-mass effect. The second term leads to an expression

$$\Delta E_a^{(1)} = -\frac{e^2}{(2\pi)^4}\int\frac{dk}{k^2}\int d\mathbf{p}'\int d\mathbf{p}\phi_a(\mathbf{p}')\{\gamma_\nu S_F(p'-k)\frac{e\gamma_\mu}{(2\pi)^3}S_F(p-k)\gamma_\nu\}\phi_a(\mathbf{p})A_\mu^e)\mathbf{p}'-\mathbf{p}) \quad \ldots\ldots(10.31)$$

with $p_0=p_0'=E_a$. Apart from a term removed by charge renormalization $\Delta E_a^{(1)}$ gives the relativistic contribution to the Lamb shift described earlier. The integral over the term in curly brackets gives essentially $\Lambda_{\mu c}^{(2)}(q)$, $q=p'-p$, after charge renormalization. In the limit of weak fields, certainly applicable in this case, $q^2/m^2 \ll 1$, the result may be given approximately as

$$\Delta E_a^{(1)} = \frac{i\alpha}{4\pi}\int d\mathbf{p}\int d\mathbf{q}\left[\phi_a(\mathbf{p}+\mathbf{q})\left\{\frac{i}{2m}(q\gamma_\mu-\gamma_\mu q)+\frac{4q^2}{3m^2}\left(\log\frac{m}{\lambda_{\min}}-\frac{3}{8}\right)\right\}\phi_a(\mathbf{p})\frac{eA_\mu^e(\mathbf{q})}{(2\pi)^3}\right] \quad \ldots\ldots(10.32)$$

where for a Coulomb field the only non-vanishing component of $A_\mu^e(\mathbf{q})$ is $A_4^e(\mathbf{q})=-iZe/|\mathbf{q}|^2$, and $\alpha=e^2/4\pi$ denotes the fine structure constant.

The remaining term in the expression for ΔE_a gives the non-relativistic contribution, and is convergent for large values of the momenta. However, it has to compensate the infra-red divergence in $\Delta E_a^{(1)}$, which has been obtained on the assumption that the photons have a small rest mass $\lambda_{\min}$, and becomes infinite as $\lambda_{\min}\to 0$. The analysis is rather complicated, and for details reference may be made to the work of Baranger, Bethe and Feynman (1953) already referred to. The decomposition of $S_F^{(e)}$ given in equation (10.30) is not in fact the most convenient to use for the separation of the non-relativistic contribution. A similar problem is dealt with by Kroll and Pollock (1952). It had been shown earlier (French and Weisskopf 1949) that if the non-relativistic contribution is put in the form given by Bethe (1947)

$$\frac{2\alpha}{3\pi m^2}\int_0^{k_{\max}} dk \sum_n |\mathbf{p}_{an}|^2\frac{E_n-E_a}{E_n+k-E_a} \quad \ldots\ldots(10.33)$$

then the fitting is to be made by taking $\log 2k_{\max}-\frac{5}{6}=\log\lambda_{\min}$.

The problem of the polarization of the vacuum has yet to be considered. The effect on the Lamb shift is a small one, and we shall only describe the perturbation treatment based on the graph of figure 7. In contrast with a term $-e\gamma_\mu A_\mu^e(q)$ contributed by the direct scattering the graph of figure 7 contributes a term which may be written $-e\gamma_\mu A_\mu^{e\prime}(q)$ where

$$A_\mu^{e\prime}(q)=\frac{ie^2}{(2\pi)^4}\frac{1}{q^2}\int dk\ \mathrm{Sp}\{\gamma_\mu S_F(k)\gamma_\nu S_F(k+q)\}A_\nu^e(q). \quad \ldots.(10.34)$$

The integral is one met earlier in § 9.3 where its general properties were described. Its evaluation (cf. Schwinger 1949 b) leads to an expression of the form

$$C(\delta_{\mu\nu}q^2-q_\mu q_\nu)-\frac{i\pi^2}{m^2}(\delta_{\mu\nu}q^2-q_\mu q_\nu)\int_0^1 du\ \frac{u^2(1-\frac{1}{3}u^2)}{1+q^2(1-u^2)/(4m^2)}\frac{q^2}{\ldots}. \quad \ldots(10.35)$$

The external potential will satisfy the Lorentz condition so that the terms in $q_\mu q_\nu$ may be dropped. The first term in (10.35) then reproduces the original potential multiplied by a constant which is in fact infinite. This must be absorbed

by charge renormalization. The matter may be looked at another way. Writing

$$A_\mu^{e\prime}(q)=\frac{1}{q^2}\,J_{\mu\nu}(q)A_\nu^e(q)$$

we see that $A_\mu^{e\prime}$ is the potential that would be produced by the current distribution described by $J_{\mu\nu}(q)A_\nu^e(q)$. We are quite at liberty either to regard this current distribution as producing a modified external field to act on our original electron, or we can regard it as a modification of the original current distribution, leading to a different interaction energy with the given external field. The points of view are quite equivalent as neither effect is observable separately, but only the interaction energy between the field and the electron.

For the remaining finite term (10.34) yields

$$A_\mu^{e\prime}(q)=\frac{\alpha q^2}{4\pi m^2}\int_0^1 \frac{u^2(1-\frac{1}{3}u^2)}{1+q^2(1-u^2)/(4m^2)}\,du A_\mu^e(\mathbf{q}) \qquad \ldots\ldots(10.36)$$

and the corresponding contribution to the Lamb shift of -27 Mc/s.

10.6. *Compton Scattering*

Corrections to order e^6 for the Compton scattering cross section have been calculated by Schafroth (1949, 1950) and Brown and Feynman (1952). The correction is a rather small one, and in its calculation no new principles are involved. The fourth order matrix element is found by adding to the second order graphs all the possible self-energy and vertex parts, so that to the appropriate order S_F, D_F, γ_μ are replaced by S_F', D_F', Γ_μ. Corresponding to each second order graph there is also one irreducible fourth order graph formed by the emission by the electron of a virtual photon before the scattering and its reabsorption after the scattering.

As might be expected the calculations show an infra-red divergence in the integral over the virtual photon momentum. The difficulty is resolved in the same way as that for the radiative corrections to elastic scattering. To the Compton cross section must be added that for double Compton scattering such that one of the two final quanta has an energy below some limit K. In this way a result independent of the infra-red cut-off $\lambda_{\min}$ is found. The calculations are unfortunately very elaborate, and there is an unresolved discrepancy between the results of the two authors. We shall give no details here, but refer rather to an interesting general result derived by Thirring (1950) that, to any order, the radiative corrections to Compton scattering vanish in the non-relativistic limit. The renormalized electron charge is thus determined by Thomson scattering. The proof is very simple. Any graph for low energy Compton scattering can be found by taking an electron self-energy graph and in two of the electron lines adding external low energy quanta. The insertion of a zero energy photon in an electron line of momentum p is essentially given by replacing $(\gamma p-im)^{-1}$ by

$$\frac{\partial}{\partial p_\mu}(\gamma p-im)^{-1}=-(\gamma p-im)^{-1}\gamma_\mu(\gamma p-im)^{-1}.$$

Summing over the insertion of two such zero energy photons anywhere along the main electron line of any self-energy graph yields

$$(\gamma p - im)\{e_\mu^{(1)} e_\nu^{(2)} \frac{\partial}{\partial p_\mu} \frac{\partial}{\partial p_\nu} S_F'(p)\}(\gamma p - im). \quad \ldots\ldots(10.37)$$

Use of the expression (9.9) for $S_F'(p)$ and choice of gauge such that $pe^{(1)} = pe^{(2)} = 0$ shows that (10.37) reduces to the uncorrected matrix element. Similar arguments can be used when the external photons interact with closed loop electrons.

§ 11. Concluding Remarks

The methods of quantum electrodynamics described above are very successful in their application to scattering problems, and to essentially equivalent problems, such as that of the Lamb shift. There are, however, three deficiencies in the treatment which it might be hoped that future developments will remove—explicit infinities still occur in the theory, no adequate treatment has been described for bound states and the solution is available only as an expansion in powers of the coupling constant e^2. The first difficulty is one of general principle, and has already been discussed in § 1. The remaining two difficulties are not important for electrodynamics, where the coupling constant is small, and there are no bound states except in an external field. The nearest phenomenon to a bound state is the positronium complex, which is unstable to the decay into two or three photons. New methods are, indeed, necessary for the treatment of Lamb shift effects in positronium. The situation is entirely different for meson–nucleon physics, where all stable nuclei form bound states of the interacting field system, and, furthermore, the coupling constant, G^2 say, is so large that no expansion in powers of it can be expected to be satisfactory. In this section we shall only describe very briefly some of the recent efforts that have been made to provide adequate treatments of the three deficiencies referred to above. No attempt is made to give a comprehensive description and the references quoted are only to be regarded as a typical selection.

Firstly with regard to the infinities of the theory it has been realized for a long time that these can, at least in part, be ascribed to the point particle picture used in the description of interacting fields. It has been proposed to remove this difficulty by the use of a non-local interaction between the fields, i.e. by the use of an interaction Lagrangian of the form $\mathcal{L}^{\mathrm{I}} = g\int \bar{\psi}(x')\,\Gamma\phi(x'')\,\psi(x''')\,F(x', x'', x''')dx''dx'''$. Such a proposal was made for example by Bloch (1952), and a meson theory of this kind is developed by Kristensen and Møller (1952). The choice of the form factor F has to be limited to preserve the causality of the theory, and in electrodynamics further difficulties arise with gauge invariance. It is shown, for example, by Chrétien and Peierls (1954) that these difficulties cannot all be overcome.

With regard to the second type of deficiency, an integral equation to describe the two-body problem was introduced by Salpeter and Bethe (1951), and derived from the basis of conventional field theory by Gell-Mann and Low (1951). Their treatments are in fact special cases of a general method described by Schwinger (1951 b) based on the properties of Green's functions of quantized fields. Gell-Mann and Low define the two-body kernel, for two Fermi particles described by Heisenberg field operators ψ and ϕ

$$K(x_1, x_2\,;\, x_3, x_4) = (\Psi_0, T[\psi(x_1)\phi(x_2)\ \bar{\psi}(x_3)\ \bar{\phi}(x_4)]\Psi_0) \qquad \ldots.(11.1)$$

where Ψ_0 is the vacuum state of the interacting fields. An integral equation can

be found for this kernel, as also for the two particle wave function described by

$$\chi(x_1, x_2) = (\Psi_0, T[\psi(x_1)\,\phi(x_2)]\Psi). \qquad \ldots\ldots(11.2)$$

It may be noted that the electron propagation function $S_F'(x-y)$ described in § 9.3 is similarly defined by

$$S_F'(x-y) = (\Psi_0, T[\psi(x)\,\bar{\psi}(y)]\Psi_0). \qquad \ldots\ldots(11.3)$$

The only applications of these integral equations that have so far been made have been by an expansion of the interaction kernel in powers of the coupling constant. In this way, for example, the level shift for positronium has been determined by Karplus and Klein (1952). The method was also used by Levy (1952) in his treatment of the two-body nuclear problem. For the determination of the Green's functions in general, Schwinger (1951 b) has given a pair of functional differential equations in terms of an arbitrary external field. Up to the present no great progress has been made in the solution of these equations by non-perturbation methods, though initial steps in this direction have been taken by Edwards and Peierls (1954).

An alternative, but related, method is the Tamm–Dancoff method. A formulation of this method that avoids vacuum difficulties has been presented by Dyson (1953). The state is now described by amplitudes of the form $(\Psi_0, C(N')\,A(N)\Psi)$ where $C(N')$, $A(N)$ are a set of creation and annihilation operators for non-interacting field states. For these amplitudes a set of coupled integral equations is found. The assumption is then made that the set can be cut off by ignoring amplitudes for more than some chosen number of particles. Some of the problems of renormalization met with in this method have been discussed by Cini (1953). The method has been extensively applied to the problem of meson–nucleon scattering where, though certain outstanding difficulties remain to be cleared (Dyson 1954), considerable progress has been achieved in securing agreement with experiment. In general the complexity of these new problems serve all the more to emphasize the simplicity of electrodynamics where the perturbation method yields a valid approximation.

References

Baranger, M., Bethe, H. A., and Feynman, R. P., 1953, *Phys. Rev.*, **92,** 482.

Barber, W. C., Becker, G. E., and Chu, E. L., 1953, *Phys. Rev.*, **89,** 950.

Bethe, H. A., 1947, *Phys. Rev.*, **72,** 339.

Bleuler, K., 1950, *Helv. Phys. Acta*, **23,** 567.

Bloch, C., 1952, *K. Danske vidensk. Selsk., Mat.-fys. Medd.*, **27,** No. 8.

Bloch, F., and Nordsieck, A., 1937, *Phys. Rev.*, **52,** 54.

Brown, L. M., and Feynman, R. P., 1952, *Phys. Rev.*, **85,** 231.

Casimir, H., 1933, *Helv. Phys. Acta*, **6,** 287.

Chrétien, M., and Peierls, R. E., 1954, *Proc. Roy. Soc.* A, **223,** 468.

Cini, M., 1953, *Nuovo Cim.*, **10,** 526, 614.

Dancoff, S. M., 1939, *Phys. Rev.*, **55,** 959.

Dirac, P. A. M., 1927, *Proc. Roy. Soc.* A, **114,** 243.

Dyson, F. J., 1949 a, *Phys. Rev.*, **75,** 486; 1949 b, *Ibid.*, **78,** 1736; 1952, *Ibid.*, **85,** 631; 1953, *Phys. Rev.*, **91,** 1543; 1954, *Proceedings of the Glasgow Conference* (London: Pergamon Press).

Edwards, S. F., and Peierls, R. E., 1954, *Proc. Roy. Soc.* A, **224,** 24.

Euler, H., 1936, *Ann. Phys., Lpz.*, **26,** 398.

Fermi, E., 1931, *Rev. Mod. Phys.*, **4,** 131.

Feynman, R. P., 1949 a, *Phys. Rev.*, **76,** 749; 1949 b, *Ibid.*, **76,** 769.

French, J. B., and Weisskopf, V. F., 1949, *Phys. Rev.*, **75,** 1240.

FUKADA, H., MIYAMOTO, Y., and TOMONAGA, S., 1949, *Progr. Theor. Phys., Osaka*, **4**, 47 and 121.
FURRY, W. H., 1937, *Phys. Rev.*, **51**, 125; 1951, *Ibid.*, **81**, 115.
GEHENIAU, J., and VILLARS, F., 1950, *Helv. Phys. Acta*, **23**, 178.
GELL-MANN, M., and LOW, F., 1951, *Phys. Rev.*, **84**, 350.
GUPTA, S., 1950, *Proc. Phys. Soc.* A, **63**, 681; 1951, *Ibid.*, **64**, 426.
HEISENBERG, W., and PAULI, W., 1929, *Z. Phys.*, **56**, 1.
HEITLER, W., 1954, *Quantum Theory of Radiation* (3rd Edn), (Oxford: Clarendon Press).
HOURIET, A., and KIND, A., 1949, *Helv. Phys. Acta*, **22**, 319.
HURST, C. A., 1952, *Proc. Camb. Phil. Soc.*, **48**, 625.
JORDAN, P., and WIGNER, E., 1928, *Z. Phys.*, **47**, 63.
KALLEN, G., 1952, *Helv. Phys. Acta*, **25**, 417; 1953, *K. Danske vidensk. Selsk., Mat.-fys. Medd.*, **27**, No. 12.
KANESAWA, S., and TOMONAGA, S., 1948, *Progr. Theor. Phys., Osaka*, **3**, 1, 101.
KARPLUS, R., and KLEIN, A., 1952, *Phys. Rev.*, **87**, 848.
KARPLUS, R., KLEIN, A., and SCHWINGER, J., 1952, *Phys. Rev.*, **86**, 288.
KARPLUS, R., and KROLL, N. M., 1950, *Phys. Rev.*, **77**, 536.
KARPLUS, R., and NEUMAN, M., 1950, *Phys. Rev.*, **80**, 380; 1951, *Ibid.*, **83**, 776.
KOBA, Z., TATI, T., and TOMONAGA, S., 1947, *Progr. Theor. Phys., Osaka*, **2**, 198.
KOENIG, S. H., PRODELL, A. G., and KUSCH, P., 1952, *Phys. Rev.*, **88**, 191.
KRAMERS, H. A., 1948, *Report of 8th Solvay Congress, Brussels.*
KRISTENSEN, P., and MØLLER, C., 1952, *K. Danske vidensk. Selsk. Mat.-fys., Medd.*, **27**, No. 7.
KROLL, N. M., and POLLOCK, F., 1952, *Phys. Rev.*, **86**, 876.
LAMB, W. E., and RETHERFORD, R. C., 1950, *Phys. Rev.*, **79**, 549.
LEVY, M., 1952, *Phys. Rev.*, **88**, 72, 725.
LEWIS, H. W., 1948, *Phys. Rev.*, **73**, 173.
MATTHEWS, P. T., 1949, *Phys. Rev.*, **76**, 684; 1950 a, *Ibid.*, **80**, 293; 1950 b, *Phil. Mag.*, **41**, 185.
MATTHEWS, P. T., and SALAM, A., 1951, *Rev. Mod. Phys.*, **23**, 311; 1954, *Phys. Rev.*, **94**, 185.
PAULI, W., 1933, *Handbuch der Physik* (2nd Edn), **24**, Pt. 1 (Berlin: Springer); 1936, *Ann. Inst. Poincaré*, **6**, 137; 1941, *Rev. Mod. Phys.*, **13**, 203; 1950, *Progr. Theor. Phys., Osaka*, **5**, 526; 1951, *Feldquantisierung*, Lectures at E.T.H., Zurich.
PAULI, W., and FIERZ, M., 1938, *Nuovo Cim.*, **15**, 167.
PAULI, W., and VILLARS, F., 1949, *Rev. Mod. Phys.*, **21**, 434.
PEIERLS, R. E., 1952, *Proc. Roy. Soc.* A, **214**, 143.
REDHEAD, M. L. G., 1953, *Proc. Roy. Soc.* A, **220**, 219.
ROHRLICH, F., and GLUCKSTERN, R. L., 1952, *Phys. Rev.*, **86**, 1.
SALAM, A., 1951 a, *Phys. Rev.*, **82**, 217; 1951 b, *Ibid.*, **84**, 426; 1952, *Ibid.*, **86**, 731.
SALPETER, E. E., 1953, *Phys. Rev.*, **89**, 92.
SALPETER, E. E., and BETHE, H. A., 1951, *Phys. Rev.*, **84**, 1232.
SCHAFROTH, M. R., 1949, *Helv. Phys. Acta*, **22**, 501; 1950, *Ibid.*, **23**, 542.
SCHWINGER, J., 1948, *Phys. Rev.*, **74**, 1439; 1949 a, *Ibid.*, **75**, 651; 1949 b, *Ibid.*, **76**, 790; 1951 a, *Ibid.*, **82**, 914; 1951 b, *Proc. Nat. Acad. Sci., Wash.*, **37**, 452, 455; 1953, *Phil. Mag.*, **44**, 1171.
STUECKELBERG, E. G. C., 1944, *Nature, Lond.*, **153**, 143.
TAKEDA, G., 1952, *Progr. Theor. Phys. Osaka*, **7**, 359.
TATI, T., and TOMONAGA, S., 1948, *Progr. Theor. Phys., Osaka*, **3**, 391.
THIRRING, W., 1950, *Phil. Mag.*, **41**, 1193; 1953, *Helv. Phys. Acta*, **26**, 33.
TOMONAGA, S., 1946, *Progr. Theor. Phys., Osaka*, **1**, 27.
TRIEBWASSER, S., DAYHOFF, E. S., and LAMB, W. E., 1953, *Phys. Rev.*, **89**, 98.
UEHLING, E. A., 1935, *Phys. Rev.*, **48**, 55.
WARD, J. C., 1950 a, *Phys. Rev.*, **77**, 293; 1950 b, *Ibid.*, **78**, 182; 1951 a, *Ibid.*, **84**, 897; 1951 b, *Proc. Phys. Soc.* A, **64**, 54.
WEISSKOPF, V. F., 1934, *Z. Phys.*, **90**, 817; 1936, *K. Danske vidensk. Selsk. Mat.-fys., Medd.*, **14**, No. 6.
WENTZEL, G., 1948, *Phys. Rev.*, **74**, 1070.
WICK, G. C., 1950, *Phys. Rev.*, **80**, 268.

MAGNETOSTRICTION AND MAGNETOMECHANICAL EFFECTS

By E. W. LEE

Department of Physics, The University, Nottingham

CONTENTS

PAGE

§ 1. Introduction 185
1.1. General 185
1.2. Early experimental work 186
1.3. Magnetostriction and magnetization 188

§ 2. Phenomenological theory of magnetostriction in crystals 192
2.1. General 192
2.2. Cubic crystal 193
2.3. Hexagonal crystals 198

§ 3. Origins of magnetostriction 198
3.1. Formal origin 198
3.2. Theoretical treatments 200

§ 4. Factors affecting the magnetostriction constants 202
4.1. Temperature 202
4.2. Composition of alloys 203
4.3. Crystallographic order in alloys 204

§ 5. Magnetostriction below saturation 205
5.1. Single crystals 205
5.2. Polycrystals 207

§ 6. Magnetostriction and the effects of stress 210

§ 7. Volume magnetostriction 214
7.1. General 214
7.2. The thermal expansion anomaly 216
7.3. Change of Curie temperature with pressure 217

§ 8. The Form effect 219

§ 9. The ΔE-effect 220

§ 10. Miscellaneous topics 225

§ 11. Conclusion 226

Acknowledgments 227

References 227

Abstract. The energy of a ferromagnetic substance comprises three parts, the exchange, anisotropy and magnetostatic (demagnetizing) energies. These all vary with the state of strain of the substance and so it will deform spontaneously if the deformation reduces the total energy (equal to the sum of the magnetic energy and the elastic energy). This phenomenon is known as magnetostriction and the three magnetic energies respectively give rise to the volume magnetostriction, the linear magnetostriction and the form effect. The greater part of this report is taken up with the linear magnetostriction since this is the effect most easily and often observed. The formal theory as developed by Becker and others is given in § 2. § 3 reviews the various attempts to calculate the magnetostriction constants in terms of interatomic forces. The fact that the linear magnetostriction can interact with stress to give rise to an additional anisotropy is discussed in § 6 together with some consequences of this fact. § 7 deals with the volume magnetostriction and related topics which arise from the variation of the exchange energy with interatomic distance. The form effect which arises incidentally in the analysis of volume magnetostriction is treated more fully in § 8. The only secondary effect of any importance, the ΔE effect, is discussed in § 9. The final section describes such topics as magnetic strain analysis and the magnetostriction of ferrites and antiferromagnetics.

§ 1. Introduction

1.1. *General*

The post war position in the field of ferromagnetism has already been well reviewed in these Reports. The first of these dealt with the fundamental aspects of intrinsic magnetization (Stoner 1948) whilst the second covered the field of technical magnetization (Stoner 1950). In the space available it was not found possible to include in those reports any detailed discussion of the numerous secondary phenomena which are exhibited by ferromagnetics as a result of their possession of spontaneous magnetization. Some of these (for example the change of thermoelectric power with magnetization, Hall effect and some aspects of the magneto-resistance effect) are as yet little understood as far as their basic origins are concerned and still require a good deal of further investigation. The most important of the secondary effects, and that most closely related to the fundamental ones, namely magnetostriction, is understood, at least in principle and has an extensive literature. This report, whilst in no way intending to supplement the above mentioned articles, attempts to review this literature and to present the present state of the field as a whole.

There are three main reasons for studying magnetostriction. In the first place such a study may throw light on the nature and behaviour of the special internal forces in a ferromagnetic crystal. Some of this information may not be obtainable by other means. Secondly, the phenomenon is of use in investigating magnetization processes and analysing magnetization curves. The third reason is purely a practical one. The essence of the magnetostriction effect is that an electrical pulse can be converted into a mechanical one and vice versa. This enables it to be used for the production of electromechanical transducers, particularly for obtaining high intensity ultrasonic waves, magnetostriction oscillators and filters of very high stability, storage delay lines etc. One of the less beneficial aspects is that a good deal of the noise produced by power transformers operating at high flux densities is caused by the magnetostrictive vibrations of the core. The information required for these various aspects of the effect are necessarily somewhat different. For the first, the magnetostriction constants alone are usually required whilst for the second the manner of the variation of magnetostriction with magnetization may be needed as well. For the third, the requirements vary with the individual application.

Partly for this reason there seems to be some uncertainty as to what precisely is meant by magnetostriction and which, of a number of manifestations of the effect, are the fundamental ones. In what follows an attempt is made to co-ordinate the various aspects of the effect. Little more than can co-ordination be attempted at this stage since detailed knowledge of the basic origins of the effect is at present lacking.

The term magnetostriction may be said to embrace all those effects involving the interaction between magnetization and mechanical stress. A ferromagnetic substance, when magnetized, is observed to exert a stress against its surroundings and as a result its dimensions will change if they are free to do so. By application of the principle of le Chatelier it may be expected that if a ferromagnetic undergoes this effect, then its state of magnetization will change if its dimensions are forcibly altered by external stresses. These two phenomena, the change of dimensions produced by a change in magnetization, and the change in magnetization with stress are both different aspects of the same thing and it is the purpose of both theoretical and experimental investigation to establish the nature of the fundamental phenomenon which gives rise to these two effects.

1.2. *Early Experimental Work*

The fact that a ferromagnetic substance changes its length when magnetized was first established by Joule (1842). Using a bar of iron and a system of mechanical levers he was able to show that iron expands along the direction of magnetization in small fields, the expansion being reduced as the field strength is increased. Joule's experiments in this field are remarkable for the thoroughness and accuracy which distinguishes so much of his work. In addition to the effect already mentioned he was able to show that although the length of an iron bar changes when magnetized, the change in volume is very much smaller in comparison. Thus the longitudinal expansion must be accompanied by a transverse contraction. This he confirmed later by direct observation. Finally, he investigated the effects of external stresses on the magnetostriction of iron and showed that, under tension, it is a contraction for all field strengths. Some of Joule's results have been recalculated and expressed in terms of current quantities; these are shown in figure 1.1. The only observation that need be made at the moment is that the effect is really quite small and corresponds to a strain of the order 10^{-5}.

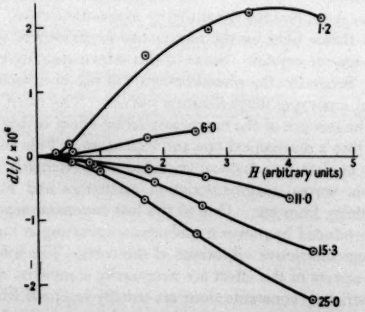


Figure 1.1. Joule's original measurements of the magnetostriction of iron (1842). The figures give the applied tension in kg mm^{-2}.

It is to be regretted that the immediately subsequent work on magnetostriction was not carried out with the insight which so characterized Joule's own. Later investigations, up to almost the end of the nineteenth century, were very largely concerned with various subsidiary and reciprocal effects which are related to the ordinary longitudinal effect. We mention the more important of these partly for the sake of completeness and partly because they still occasionally appear in the literature, but chiefly to condemn them since they tend to confuse rather than clarify the main issue. Out of a very large number of effects we may single out the following as being of greatest interest and importance.

(*a*) *Guillemin* (1846) *effect*. If a bar of ferromagnetic material is bent, either permanently or elastically, it is observed to straighten when magnetized. Since a bent bar has one half in a state of tension and the other in compression it is evident that this is merely one aspect of the influence of stress on magnetostriction and magnetization. This will be dealt with in § 6, from which it will be evident that a bent bar will always tend to straighten when magnetized, never the reverse. In addition there exists a torsional analogue, namely that if a rod or wire is given a twist then the amount of twist is decreased when it is magnetized.

(*b*) *Wiedemann* (1883) *effect*. If a ferromagnetic rod, clamped at one end is circularly magnetized by the passage of an electric current along its axis and is at the same time subjected to a longitudinal magnetic field, a twist of the free end occurs. Evidently the resultant of the two fields is a helix and the effect merely a special case of the longitudinal magnetostriction effect. This provides about the simplest means of demonstrating the existence of magnetostriction for the twist that occurs is of such magnitude that it can be observed by fixing a small mirror to the free end and noticing the deflection of a beam of light reflected from it. The effect has been very completely investigated by Pigeon (1919).

Every magnetostrictive effect has an associated inverse effect. The longitudinal magnetostriction (sometimes known as the Joule effect) has its inverse in the Villari effect (Villari 1865) which is the change of magnetization produced by an external tension. As previously mentioned these inverse effects follow from le Chatelier's principle and the Joule and Villari effects are related by the thermodynamic identity

$$\frac{1}{l}\left(\frac{dl}{dH}\right)_{\sigma}=\left(\frac{dI}{d\sigma}\right)_{H} \qquad \ldots\ldots(1.1)$$

where σ and H are the tension and magnetic field strength respectively. For the derivation of this equation and a complete discussion of the thermodynamics of magnetization, including magnetomechanical effects, reference should be made to the papers by Stoner (1935, 1937) and Guggenheim (1936). In this connection mention should also be made of a related paper by Brown (1953). The Villari effect is the most important of all the inverse effects and has been the subject of a vast amount of work (for a good account see Bozorth 1951). It will receive further attention in § 6.

The other inverse effects need not concern us. Each is governed by a thermodynamic equation with appropriate variables similar to equation (1.1). For further details the reader may be referred to an article by McKeehan (1926) and the book by Williams (1931).

Another subsidiary effect of a somewhat different nature is the change in elastic constants with magnetization. This will be treated in some detail in § 9.

1.3. *Magnetostriction and Magnetization*

The salient features of the magnetostriction effect in polycrystalline ferromagnetics may be seen from the magnetostriction–field and magnetostriction–magnetization curves of figures 1.2 and 1.3. The effect is in general quite small and varies in magnitude and sign from one material to another. It varies with the

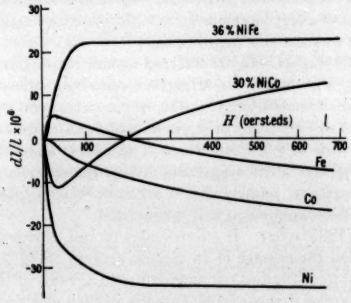


Figure 1.2. The magnetostriction of some common substances as a function of magnetic field strength.

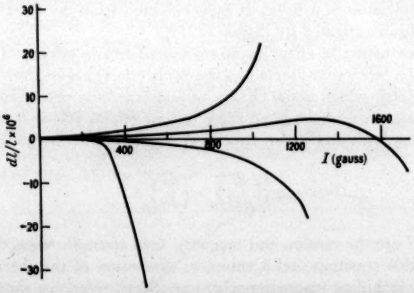


Figure 1.3. The same as in figure 1.2 as a function of the intensity of magnetization.

state of magnetization and in high fields approaches a limiting value known as the saturation magnetostriction (cf. saturation magnetization) hereafter denoted by $(dl/l)_s$. The effect exhibits hysteresis when displayed as a function of both magnetic field strength and magnetization (figures 1.4 and 1.5). In order to analyse and understand such curves, in particular those of figures 1.2 and 1.3, it is necessary to review briefly the current views on magnetization processes.

According to present day ideas a ferromagnetic substance below its Curie temperature is spontaneously magnetized to a degree which depends chiefly on its temperature and which decreases with increasing temperature. The effect of an

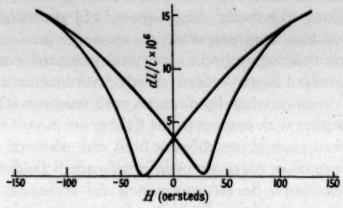


Figure 1.4. The magnetostriction of cold-worked nickel during cyclic magnetization as a function of magnetic field strength.

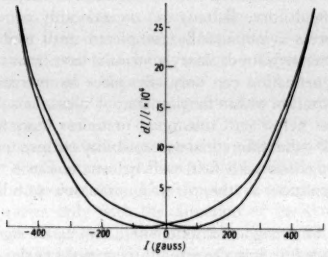


Figure 1.5. The same as in figure 1.4 as a function of the intensity of magnetization.

external magnetic field on the value of the spontaneous magnetization is usually very small. That such a material can exist in an apparently un-magnetized condition was explained by Weiss (1907) who postulated the existence of domains in the material, each of which is magnetized to saturation as stated. The direction of the spontaneous magnetization is however assumed to vary throughout the material in such a way that in the absence of an external field, the resultant component of the magnetization in any given direction is zero. In a ferromagnetic single crystal in zero applied field there is a limited number of directions which the domain magnetization can take up with respect to the crystal axes. This is known as ferromagnetic anisotropy. The number of possible directions at each point in the crystal along which the domain magnetization may lie, is governed by the symmetry of the crystal. The effects of anisotropy are usually expressed by an expression for the free energy in the form of a power series in α_1, α_2, α_3, the direction cosines of the domain magnetization vector with respect to orthogonal axes. For a cubic crystal, to which crystal class most of the ferromagnetic metals and their alloys belong, the lowest term consistent with cubic symmetry is

$$E_K = K_1(\alpha_1^2\alpha_2^2 + \alpha_2^2\alpha_3^2 + \alpha_3^2\alpha_1^2).$$

K_1 is the first anisotropy constant, a characteristic of the material, and may have either positive or negative values. If K_1 is positive E_K has a minimum value when

the domain magnetization lies along the six cube axes—[100]. If K_1 is negative the minimum occurs along the body diagonals—[111] directions. These are the so-called directions of easy magnetization.

When a magnetic field is applied to a crystal in an arbitrary direction its net magnetization is increased in two ways: firstly by movement of the boundaries separating adjacent domains whereby domains with magnetization vectors oriented favourably (energetically) with respect to the field grow in volume at the expense of their less favourably oriented neighbours. In the absence of any perturbing influences this process takes place in an infinitely small field since as long as the spontaneous magnetization—hereafter denoted I_s—is in an easy direction the anisotropy energy E_K remains unchanged. In practice this process is usually complete for ordinary materials in fields of a few oersteds. The magnetization can be further increased only by rotation of the domain vectors out of their easy directions towards the field direction. This requires the expenditure of energy, the magnitude of which is determined by the magnitude of K_1 and for most materials the magnetization–field curve flattens out considerably once this stage has been reached. This process is not usually completed until quite large fields (of the order of a thousand oersteds for ordinary materials) have been applied. Subsequent increase in the magnetization can only take place by increasing the value of the spontaneous magnetization within the domain. Unless the substance is maintained at a temperature just below its Curie point or unless huge fields are applied this increase is very small—the differential susceptibility corresponding to an increase of spontaneous magnetization with field well below the Curie temperature is of the same order of magnitude as the paramagnetic susceptibility above the Curie temperature.

If the corresponding magnetostriction–field curves are examined it can be seen that they can be divided up into the same three regions as the magnetization curves (figure 1.6). For example, in polycrystalline iron there is an initial expansion which occurs in the same field region in which the domain boundary displacements occur. This expansion is succeeded in higher fields by a gradual contraction which

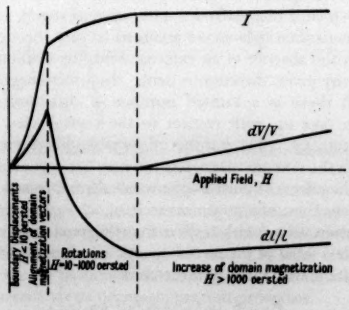


Figure 1.6. Schematic representation of magnetization, linear and volume magnetostriction of iron as a function of magnetic field strength. All ferromagnetics show essentially the same behaviour.

begins at a field strength roughly corresponding to the ' knee ' of the magnetization curve, i.e., at the field strength where the boundary displacements are more or less complete and the rotational processes are just beginning. The contraction continues until the rotations are complete and the saturation magnetization and the saturation magnetostriction are reached at the same field strength. Throughout these processes the volume of the substance remains sensibly constant. In very high fields a further expansion is found (in iron). This expansion is proportional to the field strength and differs from the low field magnetostriction in that whereas the latter takes place without change of volume, the expansion in high fields is primarily a volume effect, the expansion being the same in all directions. It is natural to associate this volume strain with the field induced increase in the spontaneous magnetization.

It is evident that we are dealing with two completely different kinds of magnetization process. The first is purely one of alignment of domain magnetization vectors whilst the second involves an increase in the spontaneous magnetization of the domain itself. These two processes give rise to different forms of magnetostriction and we shall assume that they can be treated separately. We shall now concern ourselves with magnetostriction associated with the alignment process and leave the volume magnetostriction to a later section.

In the alignment process the domain magnetization vectors change their direction; the crystallographic orientation remains fixed. It should therefore be possible to explain the magnetostriction curves by assuming that below the Curie temperature all the domains are spontaneously strained by the spontaneous magnetization within each domain. The simplest assumption is that the strain within each domain varies only with the direction of the domain magnetization with respect to a single fixed axis. We may take this to be the direction of an applied field which is coincident with the direction of observation. If the angle between the domain magnetization and this axis is θ then $dl/l=f(\theta)$. Since dl/l must remain unchanged when θ changes by 180° the simplest form for dl/l is obtained by putting $f(\theta)=A\cos^2\theta$, where A is a constant of the material. It is customary to measure the magnetostriction from a hypothetical demagnetized state for which dl/l is arbitrarily put equal to zero. This hypothetical demagnetized state is defined as one in which the domain magnetization vectors are distributed uniformly over all possible directions throughout a sphere. The mean value of $\cos^2\theta$ for such a distribution is $\frac{1}{3}$. Hence $dl/l=A(\cos^2\theta-\frac{1}{3})$. The saturation value of the magnetostriction is attained when $\theta=0$ for all the domain vectors. Thus we have, for a single domain,

$$dl/l = \tfrac{3}{2}\lambda_s(\cos^2\theta-\tfrac{1}{3}) \qquad \ldots\ldots(1.2)$$

in which λ_s is a constant representing the total strain when the substance initially in the hypothetical demagnetized state is magnetized to saturation. This implies that the magnetostriction can be described by a single constant. In particular if the direction of measurement is kept fixed and the material saturated in any direction by the application of a large field, then the variation of the magnetostriction with direction of saturation magnetization should be representable by a simple $\cos^2\theta$ curve. This seems to hold fairly well for polycrystalline samples containing no preferred orientation of crystal axes but the behaviour of single crystals is much too complicated to be described by equation (1.2) (see for example figure 1.7). Equation (1.2) has a further simple interpretation. In any material below saturation the

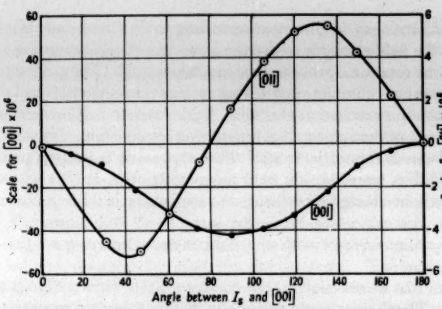


Figure 1.7. The saturation magnetostriction in the (100) plane of a single crystal of 2·52% Si–Fe, measured in the [001] and [011] directions as a function of the angle between the [001] direction and the direction of saturation magnetization (Carr and Smoluchowski 1951).

magnetization can be written $I=I_s\,\overline{\cos\theta}$, where $\overline{\cos\theta}$ is the mean value of $\cos\theta$, the projection of I_s for each domain along the direction of H (we ignore complications due to demagnetizing effects). Similarly the mean magnetostriction observed in the material is $\frac{3}{2}(\overline{\cos^2\theta}-\frac{1}{3})$. This would lead us to suppose that the magnetostriction–magnetization curves could likewise be characterized by a single constant. Inspection of figures 1.2 and 1.3 shows that over the same range of magnetization there is considerable diversity of behaviour between the various materials. Thus the likelihood of this supposition being correct is small.

We are thus led to the conclusion that the assumptions on which equation (1.2) is derived are invalid. Moreover, the fact that the behaviour of random polycrystals can be represented by equation (1.2), whereas that of single crystals in general cannot, suggests that the main error lies not in breaking off a series expansion at the first term but in taking no account of the symmetry of the crystal structure of the materials whose behaviour is to be described. This will be rectified in the next section.

§ 2. Phenomenological Theory of Magnetostriction in Crystals

In this section we show the way in which the spontaneous lattice strain must depend on the direction of the spontaneous magnetization in the lattice, and how the dependence is obtained from considerations of lattice symmetry. (N.B. Throughout this section and the next a tensor notation is used. It is to be understood that the occurrence of repeated suffixes indicates summation in the conventional manner.)

2.1. *General*

Consider a point P in a crystal lattice, whose distance from some fixed lattice point is r_0, the components of r_0 being x_0, y_0 and z_0. If this lattice is subject to a homogeneous deformation the new distance of the point P will be r with components x, y, z such that

$$x=x_0+A_{11}x_0+A_{12}y_0+A_{13}z_0$$
$$y=y_0+A_{21}x_0+A_{22}y_0+A_{23}z_0$$
$$z=z_0+A_{31}x_0+A_{32}y_0+A_{33}z_0$$

The strain tensor A_{ij} is symmetric, hence $A_{ij}=A_{ji}$. Introducing the direction cosines of the point P by the relations $\beta_1=x_0/r_0$, $\beta_2=y_0/r_0$, $\beta_3=z_0/r_0$ it is readily found that $x=r_0(\beta_i+A_{ij}\beta_j)$ with similar expressions for y and z. Hence $r^2=r_0^2(1+2A_{ij}\beta_i\beta_j)$ neglecting products of the A_{ij}, and thus $r-r_0=r_0A_{ij}\beta_i\beta_j$ so that the strain in any direction whose direction cosines are β_1, β_2, β_3 is related to the components of the strain tensor by the expression

$$dl/l=A_{ij}\beta_i\beta_j. \qquad \ldots\ldots(2.1)$$

The corresponding volume strain is $dV/V=A_{11}+A_{22}+A_{33}$.

As stated earlier it is possible to explain magnetostriction by assuming that the crystal lattice is spontaneously strained by the spontaneous magnetization, the strain depending in some way on the direction of the spontaneous magnetization with respect to the crystal axes. Thus $A_{ij}=f(\alpha_1, \alpha_2, \alpha_3)$ and the problem is to find the form of this function. It turns out that the form which f must take is completely determined by the requirements of lattice symmetry. To do this the usual procedure is to express the total energy of the crystal and then determine the equilibrium condition from the requirement that this energy should be a minimum.

This energy is made up of three parts which, following Becker and Döring, can be designated the crystal energy E_K, which is simply the crystal anisotropy energy, the magnetostriction energy E_M and the elastic energy E_L. E_K depends only on α_i whilst E_L depends only on the components of the strain tensor A_{ij}. E_M on the other hand is assumed to be a function of both A_{ik} and α_i. E_M therefore represents the interaction between the elastic and anisotropy energies. Thus it is possible to write, purely formally,

$$\begin{aligned} E_K &= a_i\alpha_i+a_{ij}\alpha_i\alpha_j+\ldots. & &+a_{ijkl}\alpha_i\alpha_j\alpha_k\alpha_l \\ E_M &= & b_{ijk}\,\alpha_iA_{jk} &+b_{ijkl}\,\alpha_i\alpha_jA_{kl} \\ E_L &= & & c_{ijkl}\,A_{ij}\,A_{kl} \end{aligned}$$

Each of these terms must satisfy the requirements of crystal symmetry since the energy is a scalar and independent of the choice of axes. The quantities a_{ij}, b_{ijkl} etc. are constants of the material and must conform to the symmetry of the crystal lattice. Thus these quantities must be invariant when subjected to the operations which characterize the symmetry properties of the crystal under investigation. The most general method for performing these operations is due to Seitz (1934) and is described in detail by Bitter (1937). It is shown that each of the constants must obey transformations of the form

$$a_{ijkl}=P_{ip}P_{jq}P_{kr}P_{ls}a_{pqrs}. \qquad \ldots\ldots(2.2)$$

The quantities P_{ip} etc. are the symmetry operators for the particular type of crystal under consideration. In general the requirements of symmetry impose severe restrictions on the number of independent constants. Fortunately the ferromagnetic metals nickel, iron and cobalt all belong to crystal classes of high symmetry and in consequence the expressions for E_K, E_M and E_L can be reduced to fairly simple forms.

2.2. *Cubic Crystals*

Crystals of iron and nickel are body- and face-centred cubic respectively and for these crystals class 0^h the symmetry operators are

$$\begin{pmatrix}1&0&0\\0&0&1\\0&-1&0\end{pmatrix} \quad \begin{pmatrix}0&0&1\\1&0&0\\0&1&0\end{pmatrix} \quad \begin{pmatrix}-1&0&0\\0&-1&0\\0&0&-1\end{pmatrix}$$

These will be referred to as I, II and III. The relations to be satisfied are those of equation (2.2) where the P's are any one of the three matrices above.

From III it follows that all terms containing an odd number of subscripts must vanish. For cubic crystals these could have been used from the start for it follows immediately from the fact that they possess a centre of symmetry. II permits cyclic interchange of subscripts $1\rightarrow 3$, $2\rightarrow 1$ and $3\rightarrow 2$, without altering the value of the term. From I it follows that the subscripts 2 and 3 may be interchanged but each interchange is to be accompanied by a change of sign. Furthermore by definition all subscripts in the E_{K} terms may be interchanged without altering their value whilst for the E_{M} and E_{L} terms i and j and also k and l may be interchanged. The energy terms become

$$E_{\mathrm{K}}=K_1(\alpha_1^2\alpha_2^2+\alpha_2^2\alpha_3^2+\alpha_3^2\alpha_1^2) \qquad \ldots\ldots(2.3)$$

dropping a constant term and replacing the constant a_{1111} by K_1. The elastic energy term contains only three independent constants and is of the form

$$\begin{aligned}E_{\mathrm{L}}=&\tfrac{1}{2}c_{11}(A_{11}^2+A_{22}^2+A_{33}^2)+c_{12}(A_{11}A_{22}+A_{22}A_{33}+A_{33}A_{11})\\&+2c_{44}(A_{12}^2+A_{13}^2+A_{23}^2). \qquad \ldots\ldots(2.4)\end{aligned}$$

The magnetostriction part E_{M} involves only two independent constants (see for example Bitter p. 175). $B_1=b_{1111}$, $2B_2=b_{1212}$. The B_1 and B_2 are then identical with the magnetoelastic coupling constants of Kittel (1949). Thus

$$\begin{aligned}E_{\mathrm{M}}=&B_1(\alpha_1^2A_{11}+\alpha_2^2A_{22}+\alpha_3^2A_{33})\\&+2B_2(\alpha_1\alpha_2A_{12}+\alpha_2\alpha_3A_{23}+\alpha_3\alpha_1A_{31}). \qquad \ldots\ldots(2.5)\end{aligned}$$

The total energy density depending on strain and on the direction of the spontaneous magnetization is the sum of these three terms. Thus

$$\begin{aligned}E=&K_1(\alpha_1^2\alpha_2^2+\alpha_2^2\alpha_3^2+\alpha_3^2\alpha_1^2)+\tfrac{1}{2}c_{11}(A_{11}^2+A_{22}^2+A_{33}^2)\\&+c_{12}(A_{11}A_{22}+A_{22}A_{33}+A_{33}A_{11})+2c_{44}(A_{12}^2+A_{13}^2+A_{23}^2)\\&+B_1(\alpha_1^2A_{11}+\alpha_2^2A_{22}+\alpha_3^2A_{33})\\&+2B_2(\alpha_1\alpha_2A_{12}+\alpha_2\alpha_3A_{23}+\alpha_3\alpha_1A_{31}). \qquad \ldots\ldots(2.6)\end{aligned}$$

The equilibrium configuration is obtained by setting $\partial E/\partial A_{ij}=0$. This gives six equations, viz.

$$\begin{aligned}\frac{\partial E}{\partial A_{11}}&=c_{11}A_{11}+c_{12}(A_{22}+A_{33})+B_1\alpha_1^2=0\\\frac{\partial E}{\partial A_{22}}&=c_{11}A_{22}+c_{12}(A_{33}+A_{11})+B_1\alpha_2^2=0\\\frac{\partial E}{\partial A_{33}}&=c_{11}A_{33}+c_{12}(A_{11}+A_{33})+B_1\alpha_3^2=0\\\frac{\partial E}{\partial A_{12}}&=4c_{44}A_{12}+2B_2\alpha_1\alpha_2=0\\\frac{\partial E}{\partial A_{13}}&=4c_{44}A_{13}+2B_2\alpha_1\alpha_3=0\\\frac{\partial E}{\partial A_{23}}&=4c_{44}A_{23}+2B_2\alpha_2\alpha_3=0. \qquad \ldots\ldots(2.7)\end{aligned}$$

The solutions are

$$A_{ii}=\frac{B_1\{c_{12}-(c_{11}+2c_{12})\alpha_i^2\}}{(c_{11}-c_{12})(c_{11}+2c_{12})}$$

$$A_{ij}=-\frac{B_2\alpha_i\alpha_j}{2c_{44}}. \qquad \ldots\ldots(2.8)$$

Now from equation (2.1)

$$dl/l=\sum A_{ii}\beta_i^2+2\sum_{i>j}A_{ij}\beta_i\beta_j.$$

Substituting the values of A_{ii} and A_{ij} given above we have

$$\begin{aligned}\frac{dl}{l}=&-\frac{B_1}{c_{11}-c_{12}}(\alpha_1^2\beta_1^2+\alpha_2^2\beta_2^2+\alpha_3^2\beta_3^2)\\&-\frac{B_2}{c_{44}}(\alpha_1\alpha_2\beta_1\beta_2+\alpha_2\alpha_3\beta_2\beta_3+\alpha_3\alpha_1\beta_3\beta_1)\\&+\frac{3c_{12}B_1}{(c_{11}-c_{12})(c_{11}+2c_{12})}. \qquad \ldots\ldots(2.9)\end{aligned}$$

By common consent for a demagnetized crystal with a perfectly random distribution of domains $dl/l=0$. In this state the mean values $\overline{\alpha_i^2}$ and $\overline{\alpha_i\alpha_j}$ are respectively equal to $\frac{1}{3}$ and 0 both when [100] and [111] are easy directions. Therefore

$$\begin{aligned}dl/l=&h_1(\alpha_1^2\beta_1^2+\alpha_2^2\beta_2^2+\alpha_3^2\beta_3^2-\tfrac{1}{3})\\&+2h_2(\alpha_1\alpha_2\beta_1\beta_2+\alpha_2\alpha_3\beta_2\beta_3+\alpha_3\alpha_1\beta_3\beta_1). \qquad \ldots\ldots(2.10)\end{aligned}$$

It may be noticed in addition that the volume change produced by the spontaneous magnetization is

$$\frac{dV}{V}=A_{11}+A_{22}+A_{33}=\frac{B_1}{c_{11}+2c_{12}} \qquad \ldots\ldots(2.11)$$

so that this volume strain is related to the magnetostriction constant h_1. It is however a good deal smaller and to this approximation is independent of α_i.

If the two-constant equation given by equation (2.10) is not adequate to describe the experimental results, then higher terms must be added. This is done by adding a term $b_{ijklmn}\alpha_i\alpha_j\alpha_k\alpha_l A_{mn}$ to E_M and proceeding as before. It turns out (Becker and Döring 1939 p. 132) that in the next approximation five constants are required and the expression is

$$\begin{aligned}dl/l=&h_1(\alpha_1^2\beta_1^2+\alpha_2^2\beta_2^2+\alpha_3^2\beta_3^2-\tfrac{1}{3})\\&+2h_2(\alpha_1\alpha_2\beta_1\beta_2+\alpha_2\alpha_3\beta_2\beta_3+\alpha_3\alpha_1\beta_3\beta_1)\\&+h_4(\alpha_1^4\beta_1^2+\alpha_2^4\beta_2^2+\alpha_3^4\beta_3^2+\tfrac{2}{3}s-\tfrac{1}{3})\\&+2h_5(\alpha_1\alpha_2\alpha_3^2\beta_1\beta_2+\alpha_2\alpha_3\alpha_1^2\beta_2\beta_3+\alpha_3\alpha_1\alpha_2^2\beta_3\beta_1)\\&+h_3s\ldots \quad \text{when } K_1>0, \text{ e.g., iron}\\&+h_3(s-\tfrac{1}{3})\ldots \text{ when } K_1<0, \text{ e.g., nickel} \qquad \ldots\ldots(2.12)\end{aligned}$$

with

$$h_1=-\frac{B_1}{c_{11}-c_{12}}, \qquad h_2=-\frac{B_2}{c_{44}}$$

$$h_4=-\frac{B_4}{c_{11}-c_{12}}, \qquad h_5=-\frac{B_5}{c_{44}} \qquad \ldots\ldots(2.13)$$

and $s=\alpha_1^2\alpha_2^2+\alpha_2^2\alpha_3^2+\alpha_3^2\alpha_1^2$.

This equation has been adjusted so that $dl/l=0$ for the demagnetized state as defined above taking into account that the mean values are as follows:

$$K_1>0, \quad \overline{\alpha_i^4}=\tfrac{1}{3}, \quad \overline{\alpha_i\alpha_j\alpha_k^2}=0, \quad \bar{s}=0$$
$$K_1<0, \quad \overline{\alpha_i^4}=\tfrac{1}{5}, \quad \overline{\alpha_i\alpha_j\alpha_k^2}=0, \quad \bar{s}=\tfrac{1}{3}.$$

The term in h_3 represents the change in volume associated with rotation of the domain magnetization vector (the constant term, given by equation (2.11), having been dropped in the zero fixing procedure). One would expect this to be small since it is zero in the lower approximation and h_3 is often put equal to zero.

For a long while the only material for which sufficient data existed to enable the constants h_1–h_5 to be evaluated was nickel for which Becker and Döring calculated the constants from the experimental results of Masiyama (1928). The agreement between experiment and theory was not very good and it was thought that further terms would be needed to describe the experimental results accurately. However it now seems that the experimental results were at fault rather than the symmetry equations and the most recent work by Bozorth and Hamming (1953) gives the data of table 2.1.

Table 2.1. Magnetostriction Constants

Material	$h_1\times10^6$	$h_2\times10^6$	$h_3\times10^6$	$h_4\times10^6$	$h_5\times10^6$
Nickel	$-68{\cdot}8\pm3{\cdot}8$	$-36{\cdot}5\pm1{\cdot}9$	$-2{\cdot}8\pm3{\cdot}8$	$-7{\cdot}5\pm5{\cdot}2$	$+7{\cdot}7\pm3{\cdot}1$
78% NiFe (quenched)	$13{\cdot}7\pm1{\cdot}0$	$2{\cdot}6\pm0{\cdot}5$	$-0{\cdot}3\pm0{\cdot}8$	$1{\cdot}1\pm1{\cdot}4$	$-0{\cdot}1\pm0\,8$
78% NiFe (slowly cooled)	$20{\cdot}9\pm0{\cdot}7$	$2{\cdot}8\pm0{\cdot}3$	$1{\cdot}7\pm0{\cdot}5$	$-1{\cdot}4\pm1{\cdot}0$	$-0{\cdot}2\pm0{\cdot}5$

It had already been stated by Goldman (1950) and by Carr and Smoluchowski (1951) that for the materials which they had studied h_4 and h_5 were negligible when h_1 and h_2 were of opposite sign. It appears from the table that this is also true when h_1 and h_2 are of the same sign. For most materials it seems likely therefore that the magnetostriction can be represented by the two-constant expression, equation (2.10), and the use of equation (2.12) can only be regarded as extravagance.

A rather more general equation than equation (2.12) has been derived in a very elegant manner by Mason (1950) who has also taken morphic effects into account. As a result of the spontaneous lattice distortion the crystal lattice can no longer be considered to obey cubic symmetry. Thus terms appear in the expression for the total energy which are zero for a cubic crystal. Effects of this type are called ' morphic effects ' (Mueller 1940) and may be taken into account formally by adding to the existing energy terms, morphic energy terms of the form $d_{ijklmn}\,\alpha_i\alpha_j\,A_{kl}\,A_{mn}$. For the ferromagnetic metals these terms are considerably smaller than the others. Such effects should be of importance in some non-metallic ferromagnetic compounds, for example cobalt ferrite ($\lambda_{100}=515\times10^{-6}$, Bozorth and Walker 1952) and a few ferromagnetic alloys, e.g., 63% FePt, ($\lambda_s=170\times10^{-6}$, Kussmann and v. Rittberg 1950) where the magnetostriction is large and the departure from normal symmetry is great.

It is sometimes convenient to know the value of the saturation magnetostriction measured from the ideal demagnetized state along certain crystallographic axes.

These can be obtained directly from equation (2.12) and for an ideal crystal are as follows:

[100], $\alpha_1=\beta_1=1,\ \alpha_2=\alpha_3=\beta_2=\beta_3=0$

$$\lambda_{100}=\tfrac{2}{3}h_1+\tfrac{2}{3}h_4 \qquad \ldots\ldots(2.14)$$

[110], $\alpha_1=\alpha_2=\beta_1=\beta_2=1/\sqrt{2},\ \alpha_3=\beta_3=0$

$$\lambda_{110}=\tfrac{1}{6}h_1+\tfrac{1}{2}h_2+\tfrac{1}{12}h_4 \qquad \ldots\ldots(2.15)$$

[111], $\alpha_i=\beta_i=1/\sqrt{3}$

$$\lambda_{111}=\tfrac{2}{3}h_2+\tfrac{2}{3}h_5. \qquad \ldots\ldots(2.16)$$

As already mentioned the two-constant equation is usually perfectly adequate to deal with the present experimental results. Putting $h_3=h_4=h_5=0$ and making use of equations (2.12), (2.14) and (2.16) we obtain the familiar form for the magnetostriction

$$dl/l=\tfrac{3}{2}\lambda_{100}(\alpha_1^2\beta_1^2+\alpha_2^2\beta_2^2+\alpha_3^2\beta_3^2-\tfrac{1}{3})+3\lambda_{111}(\alpha_1\alpha_2\beta_1\beta_2+\alpha_2\alpha_3\beta_2\beta_3+\alpha_3\alpha_1\beta_3\beta_1)$$

with $\lambda_{100}=-\frac{2}{3}\dfrac{B_1}{c_{11}-c_{12}}$, $\lambda_{111}=-\frac{1}{3}\dfrac{B_2}{c_{44}}$. $\qquad \ldots\ldots(2.17)$

This is the simplest equation consistent with anisotropic magnetostriction and contains two constants which have the significance given above, namely that λ_{100} and λ_{111} represent the total strain when a crystal is magnetized from the ideal demagnetized state to saturation along the [100] and [111] axes. Recent work (Lee 1955) suggests that this ideal demagnetized state very rarely exists in actual crystals so it is not possible to measure the saturation magnetostriction from the actual demagnetized state and assume that these values are identical with the constants λ_{100} and λ_{111}. Therefore in view of the arbitrariness associated with the above interpretation of λ_{100} and λ_{111} it is best to regard these simply as constants of the material.

If a further approximation is made, namely that $\lambda_{100}=\lambda_{111}=\lambda_s$ say, then equation (2.10) reduces to

$$dl/l=\tfrac{3}{2}\lambda_s(\cos^2\theta-\tfrac{1}{3}) \qquad \ldots\ldots(2.18)$$

where $\cos\theta=\Sigma\alpha_i\beta_i$ is the angle between the magnetization vector and the direction of observation. This expression contains no reference to the crystal axes and is therefore isotropic. The values usually assigned to λ_s are -34×10^{-6} for nickel and -7×10^{-6} for iron though in neither case does equation (2.18) fit the experimental results at all well. The two cases in which equation (2.18) may be used with justification are (i) for a polycrystal at saturation as mentioned in § 1.3, and (ii) for the individual domains—and thence to describe the magnetostriction–magnetization curves by a suitable averaging process—in a polycrystalline sample which has been cold worked so that the easy directions are determined entirely by the local strain anisotropy. These requirements are satisfied by cold-worked materials with low crystal anisotropy and large magnetostriction constants, e.g. nickel. For iron the crystal anisotropy is always greater than the strain anisotropy so that equation (2.18) cannot under any circumstances be used to describe the magnetostriction–magnetization curves.

It may be mentioned in passing that equations (2.10) and (2.12) show that the transverse magnetostriction, which is that obtained when the measuring direction

is perpendicular to the direction of magnetization, is simply a trivial special case of the longitudinal magnetostriction since it merely involves different values of β_i.

2.3. *Hexagonal Crystals*

Some ferromagnetics, e.g. cobalt and cobalt-rich alloys, form hexagonal crystals and it is of interest to know the form that the magnetostriction must take in this case. The problem is approached in exactly the same way as outlined above for cubic crystals. This has been done by Bitter (1937) and independently by Mason and Lewis (1954). Bitter's expression is (including a term in k_1 independent of α_1 and omitted by him)

$$\begin{aligned} dl/l = {} & k_1(\alpha_1^2-1)\beta_1^2+k_2(\alpha_2^2\beta_2^2+\alpha_3^2\beta_3^2) \\ & +k_3(\alpha_3^2\beta_2^2+\alpha_2^2\beta_3^2)+2(k_2-k_3)\alpha_2\alpha_3\beta_2\beta_3 \\ & +2k_4\alpha_1\beta_1(\alpha_3\beta_3+\alpha_2\beta_2) \qquad \ldots\ldots(2.19) \end{aligned}$$

where the α_i and the β_i are with respect to orthogonal axes such that Oz corresponds to the c axis and Ox or Oy to one of the hexagonal axes. The expression obtained by Mason and Lewis can be put in the same form as equation (2.19). This is the first approximation and corresponds to equation (2.10) for cubic crystals. Four constants are thus required to describe the behaviour. Mason and Lewis show that equation (2.19) is the same as that obtained for cylindrical symmetry. Terms involving α_i^4, i.e. those involving sixth order tensors, have to be employed before a characteristic hexagonal term appears. In this case eleven constants are required unless the saturation magnetization is parallel or perpendicular to the direction of measurement, in which case the number of constants reduces to eight.

The constants k_1–k_4 have been measured by Bozorth and Sherwood (1954) who obtain $k_1=-110$, $k_2=-45$, $k_3=-95$ and $k_4=-235$, all $\times 10^{-6}$. According to these authors equation (2.19) (cylindrical symmetry) is an adequate approximation for cobalt. Further details are given by Bozorth (1955).

The forms of the expressions for the magnetostriction and anisotropy energy for crystals of lower symmetry have been obtained by Mason (1954). In the following sections the discussion will be confined to cubic crystals.

§ 3. Origins of Magnetostriction

3.1. *Formal Origin*

So far the idea of a spontaneous lattice distortion has been introduced in a completely *ad hoc* manner. Moreover, the way in which the magnetocrystalline and elastic energies are supposed to interact to give the magnetostriction energy in the form stated is left unsaid. The origin of this term has been pointed out by Kittel (1949) whose treatment we now follow.

It is reasonable to suppose that the crystal anisotropy energy like the exchange energy will depend on the interatomic distance, i.e. on the state of strain of the crystal lattice. To express the strain dependence of this energy, the latter is expanded in a Taylor series, viz.

$$E_K = E_K^0 + \frac{\partial E_K}{\partial A_{ij}} A_{ij} + \tfrac{1}{2} \frac{\partial^2 E_K}{\partial A_{ij} \partial A_{kl}} A_{ij} A_{kl} + \ldots \qquad \ldots\ldots(3.1)$$

In this expression E_K^0 refers to the undistorted lattice and must satisfy cubic symmetry. The second term represents what we have in § 2.2 called E_M, represent-

ing the interaction between the magnetic anisotropy energy and strain. The two are equivalent if

$$\frac{\partial E_K}{\partial A_{ij}} A_{ij} = b_{ijkl} A_{kl}$$

which on account of symmetry was reduced to two terms only. Hence

$$\frac{\partial E_K}{\partial A_{ii}} = B_1 \alpha_i^2, \quad \frac{\partial E_K}{\partial A_{ij}} = 2B_2 \alpha_i \alpha_j$$

so that the constants B_1 and B_2 called by Kittel magnetoelastic coupling constants represent the strain gradient of the anisotropy energy. The third term in equation (3.1) represents the morphic energy terms. This gives a physical explanation for the occurrence of magnetostriction which, though implicit in the work of Van Vleck (1937) does not seem to have been stated until Kittel's formulation. The formal origin of magnetostriction lies in the fact that the anisotropy energy varies with strain (i.e. interatomic distance) and if the strain is such as to decrease, numerically, the anisotropy energy, then the lattice will strain spontaneously.

There are two great merits in this realization. In the first place any physical theory of anisotropy in terms of interatomic interactions is, in principle, automatically adequate for the magnetostriction and so the two problems are reduced to one. This is actually rather less useful than may appear at first sight since there is at present no detailed satisfactory theory of ferromagnetic anisotropy. Secondly Kittel shows that the solutions of equation (2.7) depend on the direction cosines in such a way that the energy in the equilibrium state given by equation (2.6) may be expressed in the form

$$E_K = (K^0 + \Delta K)(\alpha_1^2\alpha_2^2 + \alpha_2^2\alpha_3^2 + \alpha_3^2\alpha_1^2).$$

Here K^0 refers to the unstrained crystal whilst ΔK represents the contribution to the anisotropy caused by magnetostriction. The measured anisotropy constant $K_1 = K^0 + \Delta K$. It may be shown from equations (2.6), (2.8), (2.13)–(2.16) that

$$\Delta K = \frac{9}{4}\left\{(c_{11} - c_{12})\lambda_{100}^2 - 2c_{44}\lambda_{111}^2)\right\}. \qquad \ldots\ldots(3.2)$$

Some constants for iron and nickel are given in table 3.1.

Table 3.1

	c_{11}	c_{12}	c_{44}	λ_{100}	λ_{111}	K_1	K^0	ΔK
	dyn cm$^{-2}\times 10^{12}$			$\times 10^6$			ergs cm$^{-3}\times 10^4$	
Fe	2·41[1]	1·46[1]	1·12[1]	20·7[1]	−21·2[2]	+47[3]	47·14	−0·14
Ni	2·50[4]	1·60[4]	1·185[4]	−45·9[5]	−24·3[5]	−5·9[3]	−6·01	+0·11

[1] Kimura and Ohno (1934)
[2] Carr, W. J., quoted in Carr and Smoluchowski (1951)
[3] Bozorth, private communication
[4] Bozorth, Mason and McSkimmin (1951)
[5] Bozorth and Hamming (1953) (assuming $h_3 = h_4 = h_5 = 0$)

Thus $\Delta K/K_1 \sim 2\times 10^{-3}$ for iron and $\sim 10^{-2}$ for nickel. ΔK will be large when one of the constants (particularly λ_{111}) is numerically much greater than the other.

3.2. *Theoretical Treatments*

Having given the formal origin of the spontaneous lattice strain, it is now necessary to consider the fundamental interatomic forces responsible.

The earliest serious attempts to explain the origin of magnetostriction were made by Akulov (1928) and Becker (1930) who calculated the magnetostriction that would exist if the forces were purely magnetic in origin. Each lattice point in a ferromagnetic crystal is regarded as possessing a magnetic dipole moment, all the dipoles being parallel at the absolute zero where the calculations are intended to apply. Each dipole interacts with its neighbours, the interaction being represented classically by a scalar magnetic potential

$$V=\sum_{i,j}\frac{\mu_i\mu_j}{r_{ij}^3}\left\{1-\cos^2(\mu_i, r_{ij})\right\} \qquad \ldots\ldots(3.3)$$

where the summation is over the entire crystal. It can be shown (Kittel 1949) that such a potential does not give rise to magnetic anisotropy in the first approximation either for a body or a face-centred cubic lattice. If however the lattice is allowed to deform spontaneously there is a small contribution given by equation (3.2). To this magnetic dipole energy is added the elastic energy and the whole is minimized with respect to the strains to find the equilibrium conditions. It turns out that the state of lowest energy is that in which the crystal is distorted from its otherwise cubic form. On Becker's calculations the magnetostriction constants are

$$\lambda_{100}=\frac{2SI_s^2}{c_{11}-c_{12}}, \quad \lambda_{111}=\frac{4}{3}\frac{SI_s^2}{c_{44}} \qquad \ldots\ldots(3.4)$$

for a body-centred lattice whilst for a face-centred cubic lattice the constants are just one half these. Here I_s is the spontaneous magnetization equal to $N\mu$ where N is the number of interacting dipoles per unit volume, each of moment μ, and S is a constant depending on the geometry of the crystal lattice, being equal to 0·4 for body-centred cubic and 0·6 for face-centred cubic. Akulov's result differs somewhat from that of Becker and this was shown by Powell (1931) to be due to the fact that since V involves a summation over all lattice points in the crystal the effect of crystal shape must come into play. Powell shows that Becker's calculations apply to a spherical crystal whereas Akulov's results hold for an infinite flat sheet.

The magnetostriction and anisotropy constants due to dipolar forces calculated from equations (3.2) and (3.4) are given in table 3.2 and compared with experiment.

Table 3.2. Dipolar Magnetostriction and Anisotropy of Iron and Nickel

	$\lambda_{100}\times10^6$	$\lambda_{111}\times10^6$ calculated	$K_1\ (=\Delta K)$ $\times10^{-4}$	$\lambda_{100}\times10^6$	$\lambda_{111}\times10^6$ observed	$K_1\times10^{-4}$
Fe	4·9	−1·5	+0·004	20·7	−21·2	+47
Ni	0·15	−0·05	+0·00004	−45·9	−24·3	−5·9

It is evident that dipolar interactions can account for only a very small part of the observed anisotropy and magnetostriction of most ordinary materials. In addition the signs of K_1 and λ_{100} and λ_{111} are completely determined by the expressions (3.2) and (3.4). This is contrary to experience which indicates that these constants can have any sign whatever.

Vonsovsky (1940) has calculated the dipolar magnetostriction on a quantum mechanical basis and obtains an expression for the (isotropic) magnetostriction of the form $\lambda=\{S_1+S_2\cos^2\theta\}/C$ in which C is an elastic constant and θ has the same significance as in equation (2.18). S_1 and S_2 are constants involving the wave functions corresponding to the atoms in their appropriate quantum state and which may, in principle, have either positive or negative sign.

The greater part of the observed magnetostriction, like the anisotropy, is believed to have its origin in the spin–orbit interaction within the atoms. In ferromagnetic crystals the orbital moments are very largely quenched owing to inhomogeneous crystalline electric fields resulting in a kind of antiferromagnetic alignment of the orbital moments. The energy of orbital momentum depends not only on the orientations of the orbital magnetic vectors relative to each other but also on their orientation with respect to the line joining the two atoms. However the spin moments are coupled to the orbital moments (and hence to the crystal lattice) by the spin–orbit coupling. This viewpoint was suggested by Bloch and Gentile (1931) as a possible explanation of ferromagnetic anisotropy and was subsequently developed by Van Vleck (1937). After making rather large approximations and simplifications the latter was able to show that spin–orbit interactions of this nature simulated dipolar coupling (or quadrupolar coupling if $j=1$) in that the interaction energy is of the same form as equation (3.3) but with $\mu_i\mu_j r_{ij}^{-3}$ replaced by a constant S, say, which has its origin in the spin–orbit coupling and is between ten and one hundred times larger than $\mu_i\mu_j r_{ij}^{-3}$. Van Vleck shows that S is of the right order of magnitude to explain ferromagnetic anisotropy and, in principle, magnetostriction as well.

The subject has been treated by Vonsovsky (1940) whose calculations, based on a Heisenberg model, take into account both spin–spin and spin–orbit interaction. His final expressions may be written in the form

$$\lambda_{100}=-\frac{2}{3}\frac{I_s^2/I_\infty^2}{c_{11}-c_{12}}N\left\{A-B\left[\frac{1}{C}\left(1-e^{-C/kT}\right)-\frac{1}{4D}\left(1-e^{-D/kT}\right)\right]\right\}$$

$$\lambda_{111}=-\frac{1}{3}\frac{I_s^2/I_\infty^2}{c_{44}}N\left\{A'-B'\left[\frac{1}{C}\left(1-e^{-C/kT}\right)-\frac{1}{4D}\left(1-e^{-D/kT}\right)\right]\right\}$$

in which N is Avogadro's number and A, B, A', B', C and D are constants. By ascribing plausible values to these Vonsovsky shows that λ_{100} and λ_{111} are of the correct order of magnitude. The first term in the bracket represents the spin–spin part of the magnetostriction. This always varies with temperature as $(I_s/I_\infty)^2$, I_∞ being the spontaneous magnetization at absolute zero. This sort of temperature dependence has been found in polycrystalline nickel by Döring (1936). The B term represents the perturbing influence of spin–orbit interaction. Thus any departure from the law of the temperature dependence of magnetostriction $\lambda_s \propto I_s^2$ is to be regarded as due to the effects of spin–orbit interaction.

A rather different approach to the problem has been used by Katayama (1951) who takes as a starting point the calculations of Brooks (1940) of ferromagnetic anisotropy. This uses a collective electron approximation with ' tight-binding ' approximation wave functions, and spin–orbit coupling introduced as a perturbation. Effectively the only difference between the two calculations is calculating the perturbation energies. For the anisotropy the fourth order perturbation theory has to be employed since the perturbation energy is zero up to the third

order on account of the cubic symmetry. For the magnetostriction however the symmetry is destroyed by the strain and a second order treatment is sufficient. Katayama's final expressions for λ_{100} and λ_{111} are nevertheless quite complicated and for further details reference should be made to the original paper. He is able to show that the calculated magnetostriction constants are of the required order of magnitude and that their observed temperature dependence (complicated in the case of iron, see § 4) is not incompatible with theory.

From the foregoing survey it will be apparent that the difficulty of a proper theoretical treatment of magnetostriction lies not so much in our ignorance of its likely origin but in the mathematical complexities of the formulations that remain even after many drastic approximations have been made in order to make the problem tractable. In view of these difficulties it is perhaps scarcely surprising that no precise numerical estimates can be made. The calculated orders of magnitude are in reasonable agreement with experiment. At this stage it is futile to expect more.

§ 4. Factors Affecting the Magnetostriction Constants

The constants λ_{100} and λ_{111} are characteristics of the material and are changed only if the interactions responsible for magnetostriction are altered. This may be brought about in several ways.

4.1. *Temperature*

The most important influence on the magnetostriction constants is temperature. The importance of temperature measurements has been stated in § 3.2. Unfortunately measurements of the temperature variation of the magnetostriction constants present very great experimental difficulties and, with one exception to be referred to later, none of the existing measurements can be considered entirely satisfactory. Measurements of the saturation magnetostriction of polycrystalline nickel have been made by Döring (1936), Kirkham (1937), Dyakov (1947) and Sucksmith (1950). These are shown in figure 4.1. It is evident that there are

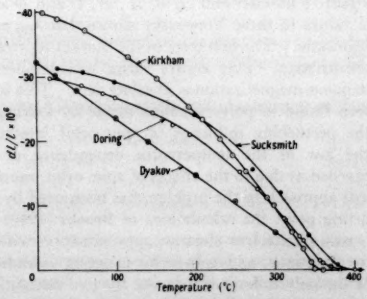


Figure 4.1. The temperature variation of the magnetostriction of polycrystalline nickel according to various observers.

considerable differences in the room temperature values and also in the actual temperature variation. This is undoubtedly caused by differences in the demagnetized state, possibly due to some preferred orientation of crystal grains. The room temperature value of Sucksmith is nearest to the accepted value and for this reason these results are probably the most reliable. The experiments of Dyakov were undertaken to provide an experimental test for the quasi-empirical relations derived by Akulov (1939) for the temperature variation of the magnetostriction, namely

$$\lambda_{100,\,T} = \lambda_{100,\,0}(1 - T/\theta) \qquad \ldots\ldots(4.1)$$

with a corresponding equation for λ_{111} and in which θ is a constant. Dyakov's experimental results obey equation (4.1) well but they differ considerably from the results of the other workers. The magnetostriction of nickel can be approximately represented by the relation $\lambda_s = CI_s^2$ which from § 3.2 presumably means that the effects of spin–orbit interaction are not very great.

For iron, measurements of λ_{100} and λ_{111} at different temperatures have been made by Takaki (1937) using as a reference state, that of the remanent magnetization. Here the variation is complicated and points to a considerable spin–orbit interaction. A 3·5% Si–Fe alloy is the only material for which satisfactory data is available. This is due to Shturkin (1947) who measured the strain at saturation as a function of crystal orientation. His results are shown in figure 4.2 together

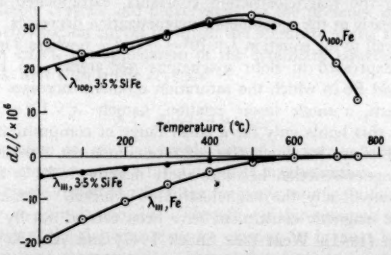


Figure 4.2. The temperature dependence of the magnetostriction constants of iron (Takaki 1936) and 3·5% Si–Fe (Shturkin 1947).

with those of Takaki for pure iron. The form of the curves for the two metals is similar, a fact which is to be expected since silicon acts simply as a diluent.

The only other materials on which any similar data exist are 15% Mo–Ni (Sucksmith 1950) and ' Alcomax ' (McCaig 1952).

4.2. *Composition of Alloys*

Another way of altering the interactions responsible for magnetostriction and thereby the magnetostriction constants is to vary the composition of alloys. The magnetostriction of binary alloys, particularly those of the ferromagnetic metals, has been extensively studied, notably by the Japanese workers, but no conclusions

seem to have been drawn from their results. The effect of alloy composition on the magnetostriction has been treated on general grounds by Vonsovsky (1940) who shows that for a binary alloy consisting of two types of atom A and B and where both A and B are magnetic, the magnetostriction should be of the form

$$\lambda_s = C\,(N_A^2 E_{AA} + 2N_A N_B E_{AB} + N_B^2 E_{BB}^2) \qquad \ldots\ldots(4.2)$$

in which C is a constant and N_A and N_B are the concentrations of atoms A and B respectively. E_{AA}, E_{BB} and E_{AB} represent the energies of interaction between like and unlike atoms. An expression exactly similar to equation (4.2) has been derived by Néel (1954a) who points out that for the nickel–iron system the magnetostriction constants can be approximately represented by

$$\begin{aligned} \lambda_{100} \times 10^6 &= -27N_A^2 + 134N_A N_B + 13N_B^2 \\ \lambda_{111} \times 10^6 &= -55N_A^2 + 340N_A N_B - 245N_B^2 \end{aligned} \qquad \ldots\ldots(4.3)$$

over the range 40–100% Ni and in which the subscript A refers to Ni. For the nickel-rich Ni–Co alloys equation (4.2) breaks down since the magnetostriction constants are linearly proportional to the cobalt concentration (Yamamoto and Miyasawa 1953).

The magnetostriction of alloy systems in which only one component is ferromagnetic has been studied, notably by Went (1951) who investigated the magnetostriction of binary alloys of nickel with a non-magnetic metal. The conclusions drawn from the investigation were: (i) In alloys containing Al, Si, Cr, V, Cu, Mo, Sa or W, the magnetostriction constants, extrapolated to 0°K, decrease numerically, linearly as the spontaneous magnetization decreases. The results can be fitted fairly well to the equation $\lambda_s \times 10^6 = -70\beta + 6$ where β is the spontaneous magnetization expressed in Bohr magnetons per atom. (ii) For alloys of Ni with Co, Mn and Fe in which the saturation moment increases with the amount of added element, a single linear relation, namely $\lambda_s \times 10^6 = 9\beta - 90$, is again found although this holds only for a small range of composition near 100% Ni. (iii) For Ni–Pd alloys in which the saturation moment remains approximately constant the magnetostriction becomes more negative. Pd is the only element that increases (numerically) the magnetostriction of nickel. Similar measurements, with iron as the magnetic constituent have been carried out by Messkin, Somin and Nekhamkin (1941), Went (see Snoek 1947) and Alizade (1950). In this case much less co-ordination is possible owing to the very great anisotropy of the magnetostriction of iron. Most of the results obtained in these investigations are so far completely unexplained.

4.3. *Crystallographic Order in Alloys*

The effect of the degree of crystallographic order on the magnetostriction of binary alloys was first studied by Goldman and Smoluchowski (1949) who showed that for a 50–50 Fe–Co alloy the magnetostriction in the ordered state was considerably greater than in the disordered state. The results were interpreted in terms of a classical dipole–dipole theory. This is unfortunate for the dipolar magnetostriction of the alloy is of the order 2×10^{-6} whereas the measured value is 92×10^{-6} in the ordered state. Thus although the theoretical analysis was confirmed by experiment it is difficult to attach much importance to the confirmation. The effect of order has been treated by Vonsovsky (1940) who points

out that if equation (4.2) represents the magnetostriction of a binary alloy, then the effect of order can be taken into account by replacing equation (4.2) by

$$\lambda_s = C\{N_A(1-pN_B)E_{AA}+2N_AN_BpE_{AB}+N_B(1-pN_A)E_{BB}\} \quad \ldots\ldots(4.4)$$

in which p is a measure of the degree of order such that $p=1$ represents complete disorder, in which case equation (4.2) is regained and $p=2$ corresponds to perfect order. Quite apart from the limited range of application of the parent equation (4.2) however, equation (4.4) does not appear to give the right sort of variation with ordering. In an A–B type alloy, the effect of ordering is to increase the number of A–B interactions, thus to increase the second term in equation (4.2) at the expense of the first and third. From equation (4.3), for the Ni–Fe system the second term is positive so that in the range of composition for which equation (4.3) holds the effect of order is to make the constants more positive, and this is contrary to the experimental results (Bozorth 1953).

It is perhaps hardly surprising that no simple expression is able to cope with changes observed during ordering. The significant fact is that the change in the magnetostriction can be very large, increases of more than 100% being observed in some cases (Goldman 1949). The existence of such changes on ordering is of great importance in the study of high permeability materials.

§ 5. Magnetostriction below Saturation

In this section we discuss the magnetostrictive behaviour of ferromagnetic materials, both single crystals and polycrystals, in the presence of applied fields, that is, the magnetostriction curves in which the literature of the subject abounds. This involves, in effect, the application of the symmetry equation (2.17). It is convenient to deal first with single crystals.

5.1. *Single Crystals*

The basic problem is, given a single crystal of known shape and size, to calculate the magnetostriction as a function of its magnetization. Equation (2.17) applies only to a single region in which the magnetization is uniform in magnitude and direction, i.e. to a single domain. Therefore if this equation is to be extended to a complete crystal the domain structure existing in it must be known for all values of the resultant magnetization. Since the domain structure will depend on the shape and size of the crystal it will be realized that this knowledge is not normally available. Even if it were, there would still remain the problem of how to calculate the mean strain in a crystal in which the various domains are deformed differently. In the general case an accurate detailed form of the magnetostriction curves ($dl/l, H$ or $dl/l, I$) is therefore hardly to be expected. Fortunately a knowledge of the domain structure becomes less important as the magnetization approaches saturation and in this case magnetostriction curves can be calculated more or less exactly.

Calculations of magnetostriction curves for crystals magnetized along the three principal crystallographic directions have been made by Akulov (1931), Heisenberg (1931), Gans and von Harlem (1933), Takagi (1939) and others. As pointed out by Becker and Döring such calculations can only be carried out unambiguously when magnetization proceeds in a known manner; this usually means when it is by domain vector rotation only. The calculations are as follows (it is

assumed that $K_1>0$; similiar considerations apply for $K_1<0$. For the time being, the detailed domain structure is ignored.

[100]. Magnetization can at any stage be by 90° or 180° boundary displacements. Hence the magnetostriction cannot be calculated as a function of the magnetization unless some assumption is made concerning the distribution of these two processes as a function of the magnetization or unless the formation can be obtained by other means. This could in principle be obtained from powder pattern investigations but so far this has not been attempted.

[110]. In low fields both 90° and 180° boundary displacements may occur and so no calculation is possible. These however are complete when all the magnetization vectors are distributed symmetrically about the two easy directions lying nearest the applied field when $I=I_s/\sqrt{2}$. Thereafter magnetization proceeds by rotations and for $I_s/\sqrt{2}<I<I_s$ the magnetostriction is

$$dl/l=\tfrac{1}{4}\lambda_{100}+\tfrac{3}{4}\lambda_{111}(2I^2/I_s^2-1). \qquad \ldots\ldots(5.1)$$

[111]. Again 90° and 180° boundary displacements can occur in low fields although the latter are the more likely. For these $dl/l=0$. Moreover from the symmetry of the direction of observation $dl/l=0$ for the 90° boundary movements as well. These are complete when $I=I_s/\sqrt{3}$ whence the magnetization can only be increased by rotations. For $I_s/\sqrt{3}<I<I_s$

$$dl/l=\tfrac{1}{2}\lambda_{111}(3I^2/I^{s2}-1). \qquad \ldots\ldots(5.2)$$

The equations (5.1) and (5.2) are shown in figure 5.1 together with some experimental points.

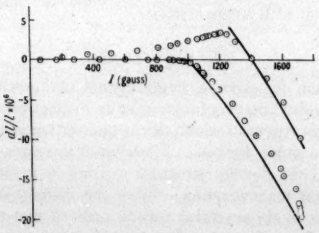


Figure 5.1. The magnetostriction of single crystals of iron in the region of domain rotation as a function of the intensity of magnetization.
Full curves calculated from equations (5.1) and (5.2) with
$I_s=1720$, $\lambda_{100}=20{\cdot}7\times10^{-6}$, $\lambda_{111}=-21{\cdot}2\times10^{-6}$
Experimental points due to Kaya and Takaki (1935).

It is also possible, in principle, to calculate magnetostriction curves in directions other than these three for crystals of suitable shapes since the basic magnetization processes are known from the work of Lawton and Stewart (1948). This has so far not been carried out however and in any case there are no reliable experimental data with which to compare the results.

Somewhat more refined calculations have been made by Lee (1955) for

magnetization along [110] using for the domain structure the model proposed by Néel (1944). For a long strip parallel to the [110] direction it is shown that the magnetostriction depends on the width of the crystal employed.

5.2. *Polycrystals*

For polycrystals the situation is still more complicated since in addition to all the difficulties mentioned in connection with single crystals there is the additional one of interaction between the crystallites. The mean magnetization in each crystal grain will in general be influenced by that of its neighbours and hence the properties of a polycrystalline aggregate will not necessarily be obtained by simply averaging the known behaviour of single crystals over all relevant orientations. In the case of magnetostriction it is by no means certain how to calculate the mean magnetostriction of such an aggregate since the strain in the isolated individual crystallites is not the same. This fact has been stressed by Vladimirsky (1943) who points out that in polycrystalline material the mechanical stresses and the elastic properties must vary from point to point in a thoroughly non-uniform manner. Vladimirksy shows that the problem is capable of an exact solution only if the shape of the crystallites and the elastic properties of the medium are known.

As mentioned in § 1.3 the saturation magnetostriction of a polycrystal varies with the direction of measurement according to $dl/l=\frac{3}{2}\lambda_s(\cos^2\theta-\frac{1}{3})$. The first problem is to calculate λ_s in terms of the single crystal constants h_1-h_5. The assumption usually made is that the stress is uniform throughout the material. It is then possible to show by averaging equation (2.12) over all possible directions that

$$\lambda_s=\frac{4}{15}h_1+\frac{2}{5}h_2+\frac{8}{35}h_4+\frac{2}{35}h_5. \qquad \ldots\ldots(5.3)$$

Using the two-constant formula, equation (2.17), it is found that

$$\lambda_s=\frac{2}{5}\lambda_{100}+\frac{3}{5}\lambda_{111} \qquad \ldots\ldots(5.4)$$

a relation first derived by Akulov (1930).

Making the assumption that the crystallites are spherical and surrounded by an infinite uniform isotropic solid, characterized by the same values of the elastic constants as the polycrystal as a whole, Vladimirksy obtains the expression

$$\lambda_s=\frac{1}{5}\frac{(c_{11}-c_{12})C_1}{\mu}\lambda_{100}+\frac{3}{5}\frac{c_{44}C_2}{\mu}\lambda_{111} \qquad \ldots\ldots(5.5)$$

where C_1 and C_2 are constants relating the strain within and outside the crystal and involve c_{11}, c_{12}, c_{44}, λ and μ, the latter two being the isotropic elastic constants (Lamé constants) as defined by Love (1936). In the case of uniform strain (an unlikely state of affairs) this expression reduces to

$$\lambda_s=\frac{1}{5}\frac{c_{11}-c_{12}}{\mu}\lambda_{100}+\frac{3}{5}\frac{c_{44}}{\mu}\lambda_{111} \qquad \ldots\ldots(5.6)$$

whilst for uniform stress equation (5.4) is obtained. In table 5.1 values of λ_s obtained from equations (5.4)–(5.6) are shown and compared with experiment. On the whole the assumption of uniform stress seems to be about the best approximation though it should be stated that λ_s is not accurately known for iron.

Table 5.1

Magnetostriction Constants (all $\times 10^6$) of Polycrystalline Iron and Nickel calculated from the Formula of Vladimirsky (eqn 5.5) and for Uniform Strain (5.6) and Uniform Stress (5.4)

	λ_{100}	λ_{111}	λ_s (5·5)	λ_s (5·6)	λ_s (5·4)	λ_s (exptl)
Fe	20·7	−21·2	− 8·9	−13·0	− 4·4	−8·0[1]
Ni	−45·9	−24·3	−31·1	−46·9	−32·9	−34[2]

[1] Kornetzki (1934) [2] Goldman (1947)

Below saturation the problem is even more complicated since the strain in the individual crystallites is not accurately known. Near saturation it is possible to express the magnetostriction as a function of the applied magnetic field strength as

$$\frac{dl}{l}=\frac{1}{5}\left(2\lambda_{100}+3\lambda_{111}\right)+\left(\lambda_{100}-\lambda_{111}\right)\left(\frac{8}{35}\frac{K}{I_s H}-\frac{32}{1001}\frac{K^2}{I_s^2H^2}\right)G$$
$$-\lambda_{111}\frac{8}{35}\frac{K^2}{I_s^2H^2}G \qquad \ldots\ldots(5.7)$$

where G is Néel's factor (1948b) for the effects of crystallite interaction. This expression, without the factor G, was first derived by Dyakov (1951) and independently by Lee (1954). The effect of magnetic interactions between the crystal grains was first correctly taken into account by Néel (1954b). Since the magnetization can be expressed in the form

$$I=I_s\left(1-\frac{8}{105}\frac{K^2}{I_s^2H^2}G\right) \qquad \ldots\ldots(5.8)$$

it is possible to obtain dl/l as a function of I but the range over which equation (5.7) is valid—approximately $0{\cdot}97I_s<I<I_s$ is too small to be of much use.

If the approximation of isotropic magnetostriction is acceptable then it is possible to obtain the magnetostriction as a function of magnetization in the following manner. Assume that at any stage of magnetization, the domain magnetization vectors are distributed uniformly within a cone of semi-angle θ_m and whose axis coincides with the direction of the applied field. Then for any domain whose magnetization vector makes an angle θ with this direction $dl/l=\frac{3}{2}\lambda_s(\cos^2\theta-\frac{1}{3})$ and $I=I_s\cos\theta$. For the whole material $\cos^2\theta$ and $\cos\theta$ must be replaced by mean values. By virtue of the assumptions made these are

$$\overline{\cos^2\theta}=\int_0^{\theta_m}\sin\theta\cos^2\theta d\theta\Big/\int_0^{\theta_m}\sin\theta d\theta$$

and

$$\overline{\cos\theta}=\int_0^{\theta_m}\sin\theta\cos\theta d\theta\Big/\int_0^{\theta_m}\sin\theta d\theta.$$

Carrying out these integrations one obtains

$$dl/l=\lambda_s\eta\,(2\eta-1) \qquad \ldots\ldots(5.9)$$

with $\eta=I/I_s$. Equation (5.19) may be expected to hold for $\eta>0{\cdot}5$ when the anisotropy is uniaxial and the magnetostriction isotropic. These conditions are

fulfilled in cold worked nickel and in hexagonal cobalt. Figure 5.2 shows that the former obeys equation (5.9) fairly well. For other materials it is found that a

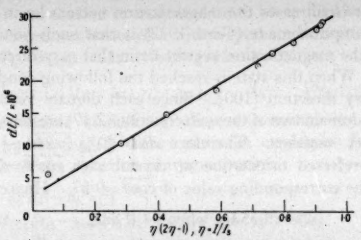


Figure 5.2. Magnetostriction of cold-worked nickel as a function of $\eta(2\eta-1)$ (see equation (5.11)).

graph of dl/l against $\eta(2\eta-1)$ gives two straight lines intersecting at a point corresponding to $\eta \sim 0{\cdot}9$, i.e. roughly where the magnetization process changes from boundary displacements to domain vector rotation (see figure 5.3). However,

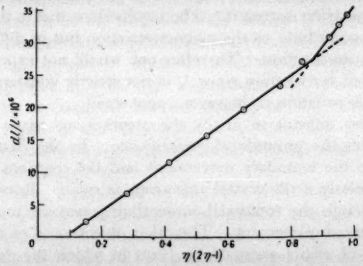


Figure 5.3. Magnetostriction of annealed nickel as a function of $\eta(2\eta-1)$.

for some substances, e.g. iron, it is not always possible to fit the (dl/l, I) curves to equation (5.9) at all.

In passing we may note that if equations (5.8) and (5.9) are combined there results the equation

$$\frac{dl}{l} = \lambda_s \left(1 - \frac{8}{35}\frac{K^2}{I_s^2 H^2} G\right) \qquad \ldots\ldots(5.10)$$

which is the isotropic form of (5.7).

If the assumption of isotropic magnetostriction is inadmissible, e.g. in the case of iron, then some fairly drastic assumptions have to be made to calculate dl/l as a function of I. In small fields magnetization proceeds by 90° and 180° boundary movements. (In nickel and for substances with $K_1 < 0$ the 90° boundaries are replaced by boundaries separating domains whose magnetization

vectors make angles of 71° and 109° with each other but this does not affect the argument and it is convenient to refer to them all as 90° boundary movements). It is therefore impossible to calculate the magnetostriction for the reasons given in § 5.1. However in all cases the magnetization vectors lie in an easy direction and the usual assumption made (Gans 1932) is that such boundary movements continue until all the magnetization vectors lie in that easy direction lying nearest the applied field. When this state is reached the following conditions hold:

(1) $K_1>0$, easy direction [100]. Since each domain vector lies in a [100] direction, for each domain two of the α_i in equation (2.17) are zero and consequently the second bracket vanishes. Therefore $dl/l=\frac{3}{2}\lambda_{100}\,(\overline{\cos^2\theta}-\frac{1}{3})$. For a polycrystal with no preferred orientation of crystal axes $\overline{\cos^2\theta}=0{\cdot}71$ (Becker and Döring 1939). The corresponding value of $\overline{\cos\theta}=0{\cdot}83$. Hence

$$dl/l=0{\cdot}55\lambda_{100} \text{ when } I=0{\cdot}83I_s. \qquad \ldots\ldots(5.11)$$

(2) $K_1<0$, easy direction [111]. Similar considerations lead to the conclusion that $dl/l=\frac{3}{2}\lambda_{111}\,(\overline{\cos^2\theta}-\frac{1}{3})$ with $\overline{\cos^2\theta}=0{\cdot}758$ (Yamamoto and Miyasawa 1953) and $\overline{\cos\theta}=0{\cdot}87$. Thus

$$dl/l=0{\cdot}637\lambda_{111} \text{ when } I=0{\cdot}87I_s. \qquad \ldots\ldots(5.12)$$

For values of I greater than these, magnetization must take place by rotations. (dl/l, I) curves have been calculated by the author (unpublished) by assuming that for all values of I each domain vector lies in the plane defined by the applied field and the easy axis lying nearest it. The results show that in the rotation region both λ_{100} and λ_{111} contribute to the magnetostriction but in different proportion depending on the magnetization. Therefore one would not expect equation (5.9) to apply exactly even for rotation since λ_s is not strictly constant. For the case in which $K_1<0$, this variation is, however, quite small.

It is nevertheless difficult to justify the assumptions made in the previous paragraph except on the grounds of convenience. In the first place it is not possible to split up the boundary movements and the rotations as two separate occurrences, particularly if the crystal anisotropy is small. Rotations will always be present even though the rotational susceptibility may be much smaller than that due to boundary displacement. The other objection is of a rather different nature. The state of affairs assumed by Gans in which the domain vectors in each crystallite all lie in a single easy direction is such as to give a non-uniform magnetization which would give rise to free magnetic poles at the surface of each crystal grain. It is most unlikely that this is allowable except near saturation. The opposite assumption, namely that the tendency to avoid free poles is so great that the mean magnetization of each crystal grain is constant in magnitude and direction, and equal to the mean magnetization of the whole material, has been developed by Lawton and Stewart (1950). Unfortunately it does not seem possible to apply their approach to the calculation of magnetostriction, unless further information is available.

§ 6. Magnetostriction and the Effects of Stress

Since the strain in a ferromagnetic crystal is determined partly by the direction of the spontaneous magnetization, it is to be expected that if such a crystal is strained by an external force, this additional strain will in general alter the direction of the spontaneous magnetization. Usually the external force is a

simple compression or tension and it will now be shown that the effect of such a stress is to introduce an additional term to the anisotropy energy superimposed on the normal crystal energy E_K.

Suppose a uniform tension of magnitude σ to be applied to a crystal such that γ_1, γ_2, γ_3 are the direction cosines of σ with respect to the crystal axes. The stress components are $T_{ij}=\sigma\gamma_i\gamma_j$. The resulting strains are:

$$A_{ii}=-\sigma\{(s_{11}-s_{12})\,\gamma_i^2+s_{12}\},\qquad A_{ij}=-\tfrac{1}{2}\sigma s_{44}\gamma_i\gamma_j$$

where the s's are elastic moduli. If these expressions are substituted in the second term of equation (3.1) there results an expression for that part of the energy depending on the direction of magnetization

$$\begin{aligned}E=&K_1(\alpha_1^2\alpha_2^2+\alpha_2^2\alpha_3^2+\alpha_3^2\alpha_1^2)\\&-B_1\sigma\{(s_{11}-s_{12})(\alpha_1^2\gamma_1^2+\alpha_2^2\gamma_2^2+\alpha_3^2\gamma_3^2)+s_{12}\}\\&-B_2\sigma s_{44}(\alpha_1\alpha_2\gamma_1\gamma_2+\alpha_2\alpha_3\gamma_2\gamma_3+\alpha_3\alpha_1\gamma_3\gamma_1).\end{aligned}$$

Using the known relations between the elastic moduli and the elastic constants and expressing B_1 and B_2 in terms of λ_{100} and λ_{111} one obtains

$$\begin{aligned}E=&K_1(\alpha_1^2\alpha_2^2+\alpha_2^2\alpha_3^2+\alpha_3^2\alpha_1^2)\\&-\tfrac{3}{2}\lambda_{100}\sigma(\alpha_1^2\gamma_1^2+\alpha_2^2\gamma_2^2+\alpha_3^2\gamma_3^2)\\&-3\lambda_{111}\sigma(\alpha_1\alpha_2\gamma_1\gamma_2+\alpha_2\alpha_3\gamma_2\gamma_3+\alpha_3\alpha_1\gamma_3\gamma_1)\qquad\ldots\ldots(6.1)\end{aligned}$$

omitting a term independent of α_i. Thus to the crystal energy is added a term $-\sigma dl/l$ usually known as the magnetoelastic energy. The values of α_i required to make E a minimum thus depend in a complicated manner on the magnitude and signs of K_1, λ_{100}, λ_{111} and σ. It may be noted that from equation (6.1) the energy depends on $\lambda_{100}\sigma$ and $\lambda_{111}\sigma$ so that a substance with positive magnetostriction constants, subject to tension, behaves exactly as one whose magnetostriction constants are negative, and submitted to compression. The effect of stress therefore is to create new energy minima for the domain vectors in addition to those given by the crystal anisotropy.

The general case, represented by equation (6.1) is extremely complicated for arbitrary values of the constants and no general solution has been put forth. The equations resulting by expressing α_i and γ_i in terms of polar coordinates θ and ϕ and setting $\partial E/\partial\theta$ and $\partial E/\partial\phi$ equal to zero are very complicated and cannot be solved explicitly for θ and ϕ. It is necessary therefore to consider approximate solutions in the two cases when the crystal energy is much greater than or much less than the magnetoelastic energy.

The simplest case occurs when the crystal energy is the dominant term in equation (6.1). It is then reasonable to assume that the domain vectors remain in one or other of the easy directions caused by the crystal energy alone; that is, that the magnetoelastic energy merely makes some of the energy minima of the crystal energy deeper than others. There are then two cases to be considered:

(1) $K_1>0$, easy direction [100]. In this case two of the α_i are zero for each domain and so the third bracket in equation (6.1) vanishes. If $\lambda_{100}\sigma$ is positive then the second term has a minimum value when each domain vector lies parallel or antiparallel to that easy direction making the smallest angle with the direction of σ. This is the case for iron under tension. If $\lambda_{100}\sigma$ is negative then it will have an (algebraic) minimum when the vectors make the largest possible angle with σ, still remaining in an easy direction.

(2) $K_1<0$, easy direction [111]. In this case $\alpha_i^2=\frac{1}{3}$ and so the term involving λ_{100} is constant, the magnetoelastic energy being determined solely by the term in λ_{111}. Here again this has an (algebraic) minimum or maximum when each domain vector lies parallel or antiparallel to the easy direction making the least or greatest angle with σ, according as $\lambda_{111}\sigma$ is positive or negative. This would occur in nickel under very small compression or tension respectively.

In each case provided the magnetoelastic energy is smaller than the crystal energy the effect of a stress σ is to make the easy axis lying nearest σ, the deepest or shallowest minimum according as $\lambda_s\sigma$ is positive or negative, where λ_s is the magnetostriction constant in the easy direction. Some consequences of this are as follows. For iron λ_{100} is positive, so that under tension, boundary movements occur in polycrystalline iron which cease when all the domain vectors are parallel or antiparallel to the easy direction lying nearest the tension. Then $I=0$ by symmetry, but $dl/l=0{\cdot}55\lambda_{100}$ with respect to the demagnetized state. For the magnetostriction is unchanged if the magnetization is reversed and so this value of dl/l is the same as that obtained in § 5.2, equation (5.11). It follows that the magnetostriction at any given field strength, under tension, is equal to the magnetostriction for the same specimen without tension, less $0{\cdot}55\lambda_{100}$ (assuming that the strain does not greatly affect the 'ease' of magnetization). This is actually observed in practice. For nickel for a similar reason the saturation magnetostriction increases in magnitude with tension but this case is complicated by the fact that for quite small stresses the magnetoelastic and crystal energies are comparable in magnitude.

When K_1 is small it is permissible to neglect the crystal energy and so the domain vector orientation is determined by the magnetoelastic energy alone. In the case in which λ_{100} is much greater or less than λ_{111} inspection of the last two terms in equation (6.1) shows that under tension this energy will be a minimum when the vectors lie parallel and antiparallel to the direction of greatest algebraic magnetostriction making the smallest angle with σ. The opposite is true for compression.

One of the consequences of the reorientation of domain vectors that occurs as a result of an external stress is that the magnetostrictive strain produced by the reorientation may actually be greater than the polycrystalline saturation magnetostriction. It follows that even though λ_{100} and λ_{111} are both positive, the saturation magnetostriction of a polycrystalline specimen under stress may be negative. This is actually observed in certain Ni–Fe alloys and was originally explained by Braunewell and Vogt (1949) in terms of a change in the magnitude and signs of λ_{100} and λ_{111} with stress. Further investigation by Elsas and Vogt (1951) showed this to be unnecessary and that all the results could be explained in terms of a reorientation process due to the applied stress.

When λ_{100} and λ_{111} are equal the magnetoelastic energy reduces to the very simple form

$$E_\sigma=-\tfrac{3}{2}\lambda_s\,\sigma\cos^2\phi \qquad \ldots\ldots(6.2)$$

in which ϕ is the angle between I_s and σ. This statement together with the assumption that $K_1=0$ is the one most often used to analyse the behaviour of ferromagnetic substances under stress. If $\lambda_s\sigma$ is positive as in 68 Permalloy under tension or nickel in compression the domain vectors align themselves parallel and antiparallel to the stress axis and the material eventually behaves like

a uniaxial crystal. The magnetization–field curve becomes rectangular and the saturation magnetostriction becomes zero. If $\lambda_s\sigma$ is negative, the stress axis becomes a difficult one and the domain vectors orientate themselves in the plane perpendicular to this. In this state, for a magnetizing–field along the stress axis, $\overline{\cos^2\theta}=\cos^2\theta=0$. At saturation $\cos^2\theta=1$ and so the saturation magnetostriction which is the difference in strain between these two states is

$$\left(\frac{dl}{l}\right)_s=\tfrac{3}{2}\lambda_s(1-\tfrac{1}{3})-\tfrac{3}{2}\lambda_s(0-\tfrac{1}{3})=\tfrac{3}{2}\lambda_s$$

or one and a half times the polycrystalline magnetostriction constant. This has been verified by Kirchner (1936) for nickel.

The total energy of the system in the presence of a magnetic field is $E=-I_sH\cos\phi-\frac{3}{2}\lambda_s\sigma\cos^2\phi$. For equilibrium $\partial E/\partial\phi=0$ whence

$$I=\frac{I_s^2}{3\lambda_s\sigma}H \quad \text{and} \quad \frac{dl}{l}=\tfrac{3}{2}\lambda_s\frac{I^2}{I_s^2} \qquad \ldots\ldots(6.3)$$

so that the magnetic susceptibility is constant and the magnetization curve degenerates into a straight line (Becker and Kersten 1930). The magnetostriction is proportional to the square of the magnetization (Kirchner 1936). From equation (6.3) it follows that the quantity $\chi_s\sigma$, χ_s being the initial susceptibility under tension large enough to cause complete reorientation, should be proportional to I_s^2/λ_s at all temperatures. From Döring's measurements $\lambda_s\propto I_s^2$ and so $\chi_s\sigma$ should be a constant independent of temperature. The experimental results of Scharff (1935) and Brockhouse (1953) are not in agreement with this prediction. The reason for this is not known.

The discussion so far has been limited to polycrystals since for these only do the experimental data exist. In the case of single crystals the situation is somewhat simplified by the fact that σ is in the same direction with respect to the crystal axes throughout. On the other hand it is now essential to take demagnetizing effects into account. A first attempt at this problem has been made by Stewart (1949) who calculated magnetization curves of single crystal Si–Fe rods subject to tension. The experimental results confirm the theoretical predictions as well as can be expected.

It has already been mentioned that, on account of symmetry, an external stress cannot, in the absence of a magnetic field, bring about the magnetization of a demagnetized material. In the presence of a field, or when the sample possesses residual magnetization, a stress can change the magnitude of the magnetization. This change is related to the magnetostriction by the thermodynamic identity

$$\left(\frac{dI}{d\sigma}\right)_H=\frac{1}{I}\left(\frac{dl}{dH}\right)_\sigma \qquad \ldots\ldots(6.4)$$

which relates the change in magnetization produced by stress to the slope of the magnetostriction–field curve. Attempts to verify equation (6.4) have, in the past, met with little success because it holds only for reversible processes and these are not easily separated from the irreversible ones. Auwers (1944) has shown that under conditions of strict reversibility and when allowance is made for demagnetizing effects equation (6.4) is obeyed within the limits of experimental error.

For a full account of the reversible and irreversible effects of stress reference should be made to the excellent account by Bozorth (1951).

The conclusions reached in this section apply only in the case of an externally applied tension. With plastically strained materials stresses may be set up inside the metal and these will vary in magnitude and direction throughout it. Thus in order to account for the observed effects it is necessary to apply equation (6.1) or (6.2) to every region where the stress may be considered uniform and to average the results. This can be done only if the stress distribution is known.

Since an external tension is not thought to affect the magnetostriction constants it is likely that these are not greatly affected by plastic strain. Thus cold-worked materials may be used in the study of magnetostriction provided that the added strain anisotropy does not make measurements too difficult.

§ 7. Volume Magnetostriction

7.1. *General*

As mentioned in the introduction, it is found that if a ferromagnetic is magnetized beyond technical saturation, the magnetostriction observed is primarily a volume change, the strain being the same in all directions. It is this volume magnetostriction to which we now turn. It is customary to study the volume change as a function of the applied field strength. The observed behaviour is usually rather complicated and will be analysed following closely the treatment of Becker (1934).

From thermodynamics it is shown that

$$\left(\frac{\partial V}{\partial H_e}\right)_p = -\left(\frac{\partial M}{\partial p}\right)_{H_e} \qquad \ldots\ldots(7.1)$$

where M is the magnetic moment, V the instantaneous volume, p the pressure and H_e the external field strength. Thus $M=VI$. If the demagnetizing factor of the specimen is N, the internal field is $H=H_e-NI$ so that $H_e=H+NI$ and $dH_e=dH+NdI$. The volume strain dV/V_0 is usually denoted by ω whence, from equation (7.1),

$$\omega=-\frac{1}{V_0}\frac{\partial}{\partial p}\int_0^{He} VI dH_e$$

$$=-\frac{1}{V_0}\frac{\partial}{\partial p}\left(\int_0^H VI dH+\tfrac{1}{2}NVI^2\right).$$

The integration and differentiation are independent and so

$$\omega=-\frac{1}{V_0}\int_0^H \frac{\partial}{\partial p}(VI)dH+\frac{1}{2}\frac{NI^2}{k} \qquad \ldots\ldots(7.2)$$

where k is the bulk modulus. It is assumed that the saturation spontaneous magnetic moment is volume dependent, the variation being expressed in terms of a quantity ν such that

$$M_s=M_s^0(1+\nu\omega) \quad \text{or} \quad VI_s=V_0I_s^0(1+\nu\omega).$$

Thus

$$\frac{\partial}{\partial p}(VI_s)=-\frac{V_0I_s^0\nu}{k}.$$

The integral in equation (7.2) is now split up into two parts, one from 0 to H_s, H_s being the least field to produce saturation, and the second from H_s to H. So that

$$\omega = \frac{1}{2}\frac{NI^2}{k} - \frac{1}{V_0}\int_0^{H_s}\frac{\partial}{\partial p}(VI)dH - \frac{I_s^0 v}{k}H_s + \frac{I_s^0 v}{k}(H_c - NI_s).$$

Thus the volume magnetostriction is made up of three parts. The first depends only on the demagnetizing factor, i.e. on the shape of the sample. It is proportional to the mean magnetization and reaches a limiting value when $I = I_s$. This is usually the dominant term in low fields. The second two terms together depend only on the saturating field H_s and not on the applied field. This part is caused by rotation of the magnetic domain vectors within the sample and is identical with the term h_3 in equation (2.12). The third part is the volume effect proper; it is proportional to the external field and is independent of the orientation of the crystal axes with respect to it (Auwers 1933). It has its origin in the volume dependence of the spontaneous magnetization. This is the fundamental effect which requires attention.

It follows that if the spontaneous saturation moment is volume dependent, then it must change with external pressure. This is related to the volume magnetostriction through equation (7.1). Before discussing the volume magnetostriction and pressure experiments it is necessary to add a word about experimental difficulties. The volume magnetostriction is very small in most ferromagnetics, the quantity $\partial\omega/\partial H$ being of the order of 10^{-10} per oersted at temperatures well below the Curie temperature. Moreover it is the isothermal value of $\partial\omega/\partial H$ which is required and this is difficult to measure since with the fields employed—of the order of a few thousand oersteds—the heating due to the magnetocaloric effect is quite large. If measurements are carried out dilatometrically some of the heat liberated from the sample may be communicated to the surrounding fluid and a spurious expansion will result. Döring (1936) sought to overcome this by using liquids of known thermal expansion and making the necessary correction. However the correction is much larger than the quantity to be measured. Moreover there still remains the thermal expansion in the sample itself which may not be allowed for in this way. This also is of the same order of magnitude as the quantity to be measured. It is evident therefore that such measurements cannot be considered satisfactory and it is significant that whereas the dilatometer measurements give $\partial\omega/\partial H = 1\times10^{-10}$ per oersted (Döring) and $0{\cdot}95\times10^{-10}$ per oersted (Snoek 1937) for nickel, the most recent direct measurements of the isothermal $\partial\omega/\partial H$ for nickel, by Azumi and Goldman (1954) give $\partial\omega/\partial H = -0{\cdot}55\times10^{-10}$ per oersted, with a possible error of 10%.

Measurements of the pressure dependence of the spontaneous magnetization present even greater difficulties and the values of $\partial M_s/\partial p$ obtained by various workers disagree violently with each other. For nickel the most recent experiments are those of Jones and Stacey (1953) which show a very large increase in the saturation moment with pressure, of the order of 10% at 10 000 atmospheres. From equation (7.1) this implies that $\partial\omega/\partial H \sim 50\times10^{-10}$ per oersted, in very poor agreement with the value obtained by direct measurement.

For iron the experiments of Ebert and Kussmann (1937) on the change of saturation moment with pressure lead to a value of $\partial\omega/\partial H$ equal to 10×10^{-10}. This is in reasonable agreement with the value 5×10^{-10} obtained directly by Kornetzki (1935).

7.2. *The Thermal Expansion Anomaly*

The volume dependence of the saturation moment also gives rise to an anomaly in the coefficient of thermal expansion near the Curie temperature, since as the temperature is increased towards the Curie point there is an additional volume change caused by the destruction of the spontaneous magnetization. The linear expansion coefficient arising from this mechanism alone is

$$\alpha_M = \frac{1}{3}\frac{\partial \omega}{\partial T} = \frac{1}{3}\frac{\partial \omega}{\partial H}\frac{\partial I_s}{\partial T}\Big/\frac{\partial I_s}{\partial H}.$$

Since $\partial I_s/\partial T$ is always negative and $\partial I_s/\partial H$ positive it follows that the sign of the expansion coefficient anomaly is always the opposite of that of $\partial\omega/\partial H$. In addition the anomaly is greatest when $\partial I_s/\partial T \div \partial I_s/\partial H$ is greatest, i.e. just below the Curie point. In nickel the anomaly is very definitely positive and this lends considerable support to the negative value of $\partial\omega/\partial H$ obtained by Azumi and Goldman.

It is of considerable importance to know the total volume change ω_s associated with the appearance of the spontaneous magnetization. From theoretical considerations Fowler and Kapitza (1929) obtained the expression

$$\omega_s = k\frac{\partial J}{\partial V}(AI_s^2 + BI_s^4) \qquad \ldots\ldots(7.3)$$

where k, A and B are constants and J is a measure of the exchange interaction. Essentially the same formula has been derived by Néel (1937). Unfortunately it is not possible to estimate ω_s from this formula since the quantity $\partial J/\partial V$, which represents the variation of exchange interaction with interatomic distance, is completely unknown. Alternatively it is possible to calculate this quantity if ω_s is known.

There are at least three ways of doing this. The most obvious is to measure the coefficient of thermal expansion over a wide enough temperature range that the ' magnetic expansion ' can be seen superimposed on the normal thermal expansion. The two can then be separated in the way shown in figure 7.1. For nickel the experimental measurements of Williams (1934) interpreted in this way give $\omega_s = -3{\cdot}24 \times 10^{-4}$ at room temperature. Of course this is a dubious procedure but one would hardly expect the estimate to be in error by more than about 50%.

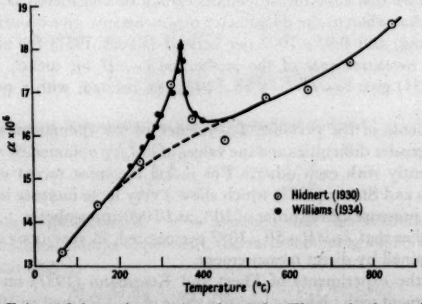


Figure 7.1. Coefficient of thermal expansion of nickel. The total volume change associated with the spontaneous magnetization is equal to the area between the full and broken curves.

The method cannot be used except for positive anomalies. For negative ones the dip in the (α, T) curve is much broader than the peak obtained for a positive anomaly and the uncertainty much greater. From x-ray measurements of the lattice constants over a wide range of temperatures by Chevenard (1927) it can be deduced that $\omega_s = +4 \times 10^{-4}$ for nickel at absolute zero (Néel 1939). This may indicate that the volume magnetostriction changes its sign at low temperatures. This could only result from a change of sign in $\partial J/\partial V$ at low temperatures, a not unlikely occurrence in view of the position usually accorded to nickel on the exchange interaction curve, namely just to the right of the maximum (see figure (7.2)). A more elegant method is due to Köster and Schmidt (1934). This involves measurements of the lattice constants of binary alloys of one of the ferromagnetic metals with a non-ferromagnetic one. For example if Mo is added to Ni, the resulting alloy system is homogeneous up to 18% Mo. Addition of Mo to Ni lowers the Curie point until it is below room temperature at about 10% Mo and the measured lattice constants show a marked change at that composition at which the alloy is no longer magnetic at room temperature. The lattice constants above this composition in the paramagnetic range vary linearly with composition and can be extrapolated back to 100% Ni to obtain the lattice constant that Ni would have if it were non-magnetic. In this way Köster and Schmidt found $\omega_s = +0{\cdot}01$ for nickel, the lattice constants in the ferromagnetic and (extrapolated) paramagnetic states being 3·515Å and 3·505Å respectively at room temperature. The third method is to use the volume magnetostriction data. According to Néel (1939) the expression $\omega_s = kI_s^2$ is a good approximation to equation (7.3). Thus $\partial\omega/\partial H = 2kI_s\partial I_s/\partial H$. Elimination of k gives

$$\omega_s = \frac{I_s}{2}\frac{\partial\omega}{\partial H}\Big/\frac{\partial I_s}{\partial H}. \qquad \ldots\ldots(7.4)$$

The chief difficulty here lies in our ignorance of $\partial I_s/\partial H$. According to Kapitza (1931) this is less than 10^{-7} for nickel. The value obtained by Polley (1939) is $1{\cdot}3 \times 10^{-4}$ and this is in good agreement with the value $1{\cdot}2 \times 10^{-4}$ estimated from theoretical considerations by Holstein and Primakoff (1940). Using this with the value $-0{\cdot}55 \times 10^{-10}$ for $\partial\omega/\partial H$, equation (7.4) gives $\omega_s = -1 \times 10^{-4}$ in fair agreement with the estimate from thermal expansion. To summarize, one can say that for nickel the volume magnetostriction and expansion anomaly data are at least in qualitative agreement. The value of ω_s deduced by Köster and Schmidt is incompatible with these both in magnitude and sign. Part of the disagreement is no doubt due to the smallness of the effects which itself may be due to the fact that the position of nickel on the exchange interaction curve is just to the right of the maximum at room temperature. Data for other metals are less complete but the agreement for nickel does not inspire confidence. There is considerable need for further experimental work on this rather untidy topic. It should be mentioned that the greater part of the theoretical work on this subject is due to Néel (1937, 1939), who has used the available data to construct curves showing the variation of exchange interaction with interatomic distance. For further information reference should be made to these two papers.

7.3. *Change of Curie Temperature with Pressure*

It is somewhat easier to investigate experimentally the pressure dependence of the Curie temperature than that of the saturation moment, and as a result the data

for the former are much more complete than for the latter. The Curie temperature is a measure of the exchange interaction and since this is known to depend on interatomic distance, it is bound to be indirectly pressure dependent. The present status of the theory of ferromagnetism is such that any direct calculation of the change of Curie temperature with pressure is out of the question. Instead the theoretical treatments have been to use thermodynamic arguments to derive relations between measured quantities and also to obtain information concerning quantities which cannot be measured directly.

The earliest attempt at a correlation process of this kind was made by Kornetzki (1935) who used a molecular field treatment, and, by thermodynamic reasoning, related the pressure dependence of the Curie temperature to the volume magnetostriction by means of the equation

$$\frac{1}{\theta}\frac{\partial\theta}{\partial p} = \frac{1}{T}\frac{\partial\omega/\partial H}{(\partial M_s/\partial T)_p - 3\alpha k\partial\omega/\partial H} \qquad \ldots\ldots(7.5)$$

in which θ is the Curie temperature, α the coefficient of thermal expansion, and k the bulk modulus. Usually the second term in the denominator is much smaller than the first and since $(\partial M_s/\partial T)_p$ is always negative, a positive $\partial\theta/\partial p$ is associated with a negative $\partial\omega/\partial H$ and vice versa.

The exchange interaction, or molecular field coefficient, in ferromagnetics is usually represented by the well-known curve due to Slater (1930) and Bethe (1933) shown in figure 7.2. Now the Curie temperature is proportional to the magnitude

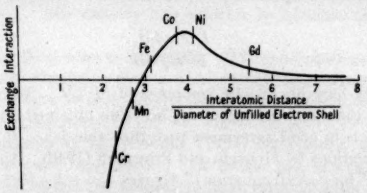


Figure 7.2. Bethe–Slater type interaction curve.

of the exchange interaction whatever particular model is used, and hence one would expect at least a qualitative correlation between the Curie temperature and the value of the ordinate for the given metals on the curve. For nickel we have seen that the most reliable measurements indicate that $\partial\omega/\partial H$ is negative and so $\partial\theta/\partial p$ should be positive, i.e. nickel should be placed to the right of the maximum in figure 7.2. This is in qualitative agreement with the most recent measurements of Patrick (1954). For other metals the predictions of figure 7.2 are in complete contradiction to the results obtained by Patrick who finds $\partial\theta/\partial p$ to be zero for iron and cobalt and what is worse, to be negative for gadolinium. Patrick suggests that the interaction curve of figure 7.2 is valid only for a single element. That the form of the curve is essentially correct seems to be confirmed by the fact that while the experimental results show quite large decreases of Curie point with pressure (e.g. for 30% Ni–Fe) there have been no large increases found. Similar conclusions can be drawn from the volume magnetostriction and expansion anomaly data.

Equation (7.5) was derived by Kornetzki on the assumption that the spontaneous

magnetization is volume independent. The particular advantage of this assumption is that it enabled him to avoid a specification of the exact form of the temperature dependence of the spontaneous magnetization. This restriction has been removed by Smoluchowski (1941 a, b) who, still using a molecular field treatment, assumed that at constant volume this dependence could be represented by a Brillouin function. Smoluchowski finds that the experimental data can be interpreted in this way only if $j=1$ for both iron and nickel. For $j=\frac{1}{2}$ the agreement between theory and experiment is poor, in some cases even disagreeing as to sign. This is interpreted as evidence for coupled spins in nickel in agreement with the conclusions of Mott and Jones (1936). Unfortunately the very fact that the theoretical results depend so sensitively on the form of the (I_s, T) curve is in itself a warning not to place too much reliance on the results obtained by using a special form of the relation $I_s=f(T)$. The spontaneous magnetization (even at constant volume) of iron and nickel cannot in all probability be accurately described by a Brillouin function and it may be suggested that the predictions of the theory would be completely different if the correct form of $I_s=f(T)$ were used. Nevertheless Smoluchowski is certainly right in removing the restriction of a volume independent magnetization. On the simplest picture of overlapping bands this clearly cannot be valid. It is unfortunate that the more general treatment cannot be viewed with complete confidence.

Some of the effects produced by hydrostatic pressure are quite appreciable. One consequence of this is that one would expect similar though smaller effects and of the opposite sign to be produced by plastic strain. There is increasing evidence (Averbach and Warren 1952) that a heavily strained material contains a high density of dislocations. In the immediate vicinity of a dislocation one would expect both the Curie temperature and the spontaneous magnetization to vary considerably on account of the large local variations in lattice constant. Experiments to show the existence of such changes have so far proved inconclusive.

§ 8. The Form Effect

Even if the spontaneous lattice deformations produced by both the exchange and anisotropy energies were zero, ferromagnetic materials would still exhibit a magnetostriction effect owing to the fact that their demagnetizing energy is dependent on strain. For example a sphere of ferromagnetic material will tend to decrease in volume and increase in length along the direction of magnetization since both changes decrease the demagnetizing energy, the latter by lowering the demagnetization factor. This effect gives rise to an additional magnetostriction known as the form effect. It has been previously encountered in connection with the volume magnetostriction. The magnitude of the effect can be calculated classically as follows.

Consider a ferromagnetic ellipsoid of revolution whose demagnetizing factor along the long axis is N. Choose orthogonal axes such that Ox coincides with this long axis. The demagnetizing energy is $\frac{1}{2}NI^2$ per unit volume in the unstrained state. If allowed to deform this energy will be

$$E_D=\tfrac{1}{2}NI^2(1-\omega)=\tfrac{1}{2}NI^2(1-A_{11}-A_{22}-A_{33})$$

using the same notation as in previous sections. The elastic energy of an isotropic solid is

$$E_L=\tfrac{1}{2}c_{11}(A_{11}^2+A_{22}^2+A_{33}^2)+c_{12}(A_{11}A_{22}+A_{22}A_{33}+A_{33}A_{11})$$
$$+(c_{11}-c_{12})(A_{12}^2+A_{13}^2+A_{23}^2).$$

The total energy $E=E_D+E_L$ and the conditions of equilibrium are $\partial E/\partial A_{ij}=0$. This gives (neglecting second order terms)

$$\frac{\partial E}{\partial A_{11}}=\tfrac{1}{2}NI^2+\tfrac{1}{2}I^2\frac{\partial N}{\partial A_{11}}-c_{11}A_{11}-c_{12}(A_{22}+A_{33})=0 \quad \ldots\ldots(8.1)$$

$$\frac{\partial E}{\partial A_{22}}=\tfrac{1}{2}NI^2+\tfrac{1}{2}I^2\frac{\partial N}{\partial A_{22}}-c_{11}A_{22}-c_{12}(A_{33}+A_{11})=0 \quad \ldots\ldots(8.2)$$

$$\frac{\partial E}{\partial A_{33}}=\tfrac{1}{2}NI^2+\tfrac{1}{2}I^2\frac{\partial N}{\partial A_{33}}-c_{11}A_{33}-c_{12}(A_{11}+A_{33})=0. \quad \ldots\ldots(8.3)$$

For an ellipsoid of revolution it is permissible to write

$$N=N_0[1-a\{A_{11}-\tfrac{1}{2}(A_{22}+A_{33})\}]$$

for N must remain unchanged by a uniform dilation $A_{11}=A_{22}=A_{33}$. Thus

$$\frac{\partial N}{\partial A_{11}}=-N_0a, \quad \frac{\partial N}{\partial A_{22}}=\frac{\partial N}{\partial A_{33}}=\tfrac{1}{2}N_0a.$$

Adding equations (8.1)–(8.3) and expressing the elastic constants c_{11} and c_{12} in terms of the more usual bulk and shear moduli k and n, one obtains for the total volume strain caused by the form effect:

$$\omega_f=\tfrac{1}{2}NI^2/k.$$

From equation (8.1) the longitudinal magnetostriction due to the form effect is

$$\left(\frac{dl}{l}\right)_f=A_{11}=\tfrac{1}{2}NI^2\left(\frac{1}{3k}+\frac{a}{2n}\right) \quad \ldots\ldots(8.4)$$

and from equation (8.2) the accompanying transverse contraction is

$$\left(\frac{dl}{l}\right)_{f,\,\text{trans.}}=A_{22}=A_{33}=\tfrac{1}{2}NI^2\left(\frac{1}{3k}-\frac{a}{4n}\right).$$

These results were first obtained by Becker (1934). Equation (8.4) has been verified experimentally by Kornetzki (1934). For an iron sphere $(dl/l)_f\sim 4\times 10^{-6}$.

The analysis given here holds only for isotropic solids. The calculation of the effect in single crystals has been carried out by Carr and Smoluchowski (1951). In general the contribution due to the form effect is small. However, with the recent tendency to use discs or oblate spheroids for experimental work, a knowledge of the part played by the shape of the sample is essential for accurate evaluation of the magnetostriction constants.

§ 9. The ΔE-effect

It has been shown in § 6 that when a ferromagnetic substance is subjected to a longitudinal tension, a redistribution of the domain magnetization vectors occurs, even in the absence of a magnetic field. This redistribution will in general be accompanied by a magnetostrictive strain ϵ_m, of magnitude $\frac{3}{2}\lambda_s\overline{(\cos^2\theta)}_\sigma-\frac{3}{2}\lambda_s\overline{(\cos^2\theta)}_{\sigma=0}$ in an obvious notation. The total strain produced by the tension is thus the sum of the ordinary elastic strain ϵ_0 and the magnetostrictive strain ϵ_m. Now ϵ_m will be a function of the applied stress and as the latter is increased ϵ_m will approach a limiting value ϵ_{ms} when the domain vectors are completely aligned by the tension in a manner given in § 6. It follows that for a ferromagnetic substance with non-

zero magnetostriction the measured stress–strain relation will be non-linear until the stress attains a value sufficiently great to cause the redistribution process to be complete. It is evident that if this redistribution is prevented from taking place, for example, by the internal magnetic anisotropy or by application of a magnetic field sufficiently great to produce saturation, the ratio stress to strain, i.e. Young's modulus, will be different from that obtained when the redistribution is allowed. This means that the measured Young's modulus of a ferromagnetic will be a function of its magnetization. This phenomenon first investigated in detail by Honda and Terada (1907) is usually known as the ΔE-effect. The usual measure of the magnitude of the effect is the difference in the values of Young's modulus in the demagnetized and magnetically saturated conditions. For annealed nickel this difference is of the order of several per cent; for iron, rather less than 1%.

Attention has been drawn in § 6 to the similarity between the effects produced by a small tension and a small magnetic field in promoting boundary movements. In both cases these will be only partly reversible. It follows that the stress–strain curve of a ferromagnetic substance will exhibit hysteresis even for stresses which are so small that the ordinary elastic hysteresis is negligible. This has been observed by Becker and Kornetzki (1934) in nickel wires subject to torsion.

It is well known that in weak magnetic fields the magnetization obeys Rayleigh's law, viz.

$$I=aH+bH^2$$

where a and b are constants. Arguing that a tension and a magnetic field act in a very similar manner Kornetzki (1938) has suggested that the magnetic strain depends on the applied stress according to a quadratic law of the same form as Rayleigh's, i.e.

$$\epsilon_m=\frac{\sigma}{E'}+\frac{\sigma^2}{E''}. \qquad \ldots\ldots(9.1)$$

The magnetic strain ϵ_m is superimposed on the normal elastic strain ϵ_0 equal to σ/E_0. Thus the total strain

$$\epsilon=\epsilon_0+\epsilon_m=\sigma\left(\frac{1}{E_0}+\frac{1}{E'}+\frac{\sigma}{E''}\right). \qquad \ldots\ldots(9.2)$$

By analogy with Rayleigh's law the remanent strain when the stress is removed is

$$\epsilon_{mr}=\sigma^2/2E''.$$

The value of Young's modulus in the demagnetized state is given by

$$\frac{1}{E_a}=\left(\frac{\epsilon}{\sigma}\right)_{\sigma\to 0}=\frac{1}{E_0}+\frac{1}{E'}.$$

At saturation $\epsilon_m=0$ and so the value of Young's modulus is simply E_0. From equation (9.2) it follows that Young's modulus defined as σ/ϵ is stress dependent and it is easy to show that

$$E=E_a\left(1-\frac{E_a}{E''}\sigma\right)$$

for small stresses. A linear decrease of E with strain has been observed by Forster and Köster (1937). This decrease is fortunately small ($E^{-1}dE/d\epsilon\sim 5\times 10^{-6}$ for Ni) compared with the ΔE effect as usually defined; otherwise it would be necessary to specify the strain (or stress) when measuring Young's modulus in the demagnetized state.

If equation (9.2) is valid then the stress–strain curve has the form shown in figure 9.1. From this it follows that the instantaneous value of Young's modulus

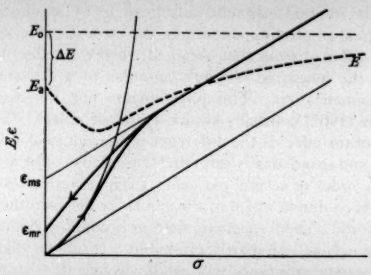


Figure 9.1. Idealized stress–strain diagram of a magnetically soft ferromagnetic substance according to Kornetzki (1938), eqn 8.

$E=\sigma/\epsilon$ decreases at first with increasing σ, goes through a minimum value and finally approaches the value E_0. This is observed in practice. It thus appears that equation (9.2) is capable of giving a good representation of the experimental facts. Nevertheless it should be borne in mind that equation (9.1) cannot be derived from Rayleigh's law and has no theoretical basis.

The customary measure of the magnitude of the ΔE-effect is the difference between the value of Young's modulus in the demagnetized state when the degree of redistribution of domain vectors is determined solely by the internal anisotropy, E_a, and its value at saturation when no such redistribution is possible, E_0. This difference will depend on (1) the distribution of domain vectors in the demagnetized state since the influence of the tension is determined by the mean value of $\cos^2(\sigma, I_s)$, (2) the magnetostriction constants since these determine the magnitude of the magnetic strain ϵ_m, and (3) on the ratio of the magnetoelastic energy to the internal anisotropy energy. If for example the crystal energy or the internal strain energy is large compared with the magneto-elastic energy associated with the external tension, then the equilibrium directions of I_s will be determined chiefly by the former; no reorientation can take place in the demagnetized state and so E_0–E_a is very small. Since the magnetoelastic energy involves the magnetostriction constants it follows that E_0–E_a will be of the form $a\lambda_s^2/C$ where a is a constant involving the geometry of the demagnetized state and C is a general internal anisotropy constant. Moreover from the conclusions reached in § 6 it is evident that ϵ_m is always positive whatever the sign of λ_s and so E_0–E_a is always positive. It should be stressed that E_0, the value at saturation, is the normal Young's modulus, i.e. the value it would possess if the magnetostriction were zero. The smaller value of E in the demagnetized state can be regarded as the effect of the magnetostrictive expansion ϵ_m, giving rise to an additional negative Young's modulus of magnitude σ/ϵ_m.

That these predictions are fairly well confirmed by experiment may be seen from figure 9.2. In the first place E_0–E_a has a maximum around 200°C. This coincides exactly with the position of the maximum initial permeability found by Kirkham

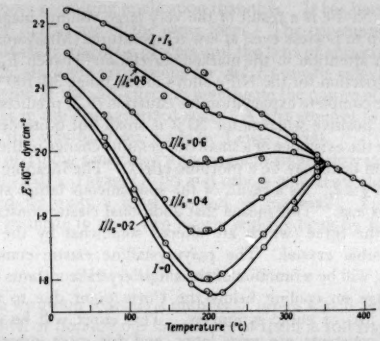


Figure 9.2. Temperature variation of Young's modulus of polycrystalline nickel for different values of magnetization (Siegel and Quimby 1936).

(1937) and is presumably associated with a minimum in C. Secondly the (E, T) curve for $I=I_s$ is more or less continuous with that for paramagnetic nickel above the Curie point. We shall refer to this later. For some materials, notably Ni–Fe alloys containing about 40% Ni, the (E, T) curve at saturation is very much below the (E, T) curve extrapolated back from the paramagnetic region (see figure 9.3).

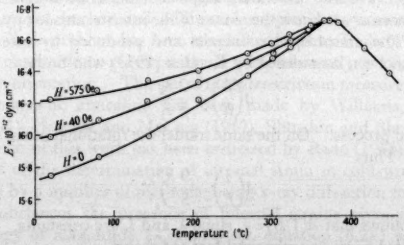


Figure 9.3. Temperature variation of Young's modulus of 42% Ni–Fe for different magnetizations showing marked change in dE/dT at the Curie temperature (Engler 1938).

This is quite contrary to expectation. Döring (1938) has proposed an explanation along the following lines. Above the Curie point one measures E at constant magnetization E_I, below it, E at constant field strength, E_H. From thermodynamics it can be shown that

$$\frac{1}{E_H}-\frac{1}{E_I}=\left(\frac{\partial \epsilon}{\partial H}\right)^2 \Big/ \frac{\partial I}{\partial H}. \qquad \ldots\ldots(9.3)$$

It is suggested that $\partial\epsilon/\partial H$ arises from the volume magnetostriction; thus $\partial\epsilon/\partial H=\frac{1}{3}\partial\omega/\partial H$. If $\partial\omega/\partial H$ is large the change in E at the Curie point may be appreciable, and Döring has suggested that the sudden change in $\partial E/\partial T$ at the Curie point,

observed in 42% Ni–Fe is a result of the very large volume magnetostriction that this alloy is known to possess even at low temperatures (Masiyama 1931). Köster (1943) has drawn attention to the marked correlation between E_H and E_I and the volume magnetostriction for the NiFe alloys in the range 30–100% Ni. However this cannot be the complete explanation for equation (9.3) predicts that $1/E_H - 1/E_I$ should always be positive whereas for Ni it is small but quite definitely negative. This may indicate the existence of a small but genuine change in the elastic constants with magnetization or it may be a morphic effect. The meaning of this term has been explained in § 2.2. As a result of the spontaneous lattice strain the original cubic symmetry is lost. This means that additional elastic constants are required over and above the three (which are altered somewhat by the morphic effect) required for a cubic crystal. The polycrystalline elastic constants, including Young's modulus, will be a function of the single crystal constants and thus may be expected to change on cooling below the Curie point due to the spontaneous deformation and loss of cubic symmetry. This effect will be small unless the magnetostriction constants are very large, and for most substances would be completely masked by the decrease in E caused by the volume magnetostriction.

A theoretical calculation of the magnitude of the ΔE-effect is possible in certain cases. For this purpose the quantity $\Delta(1/E)$ i.e. the difference in the reciprocal Young's modulus at saturation and in the demagnetized state is the most useful.

$$\Delta\left(\frac{1}{E}\right)=\frac{1}{E_s}-\frac{1}{E_0}=\frac{\epsilon_m}{\sigma}=\frac{1}{l}\left(\frac{dl}{d\sigma}\right)_{\sigma\to 0}$$

is therefore simply the magneostrictive strain produced by the stress. The effect of an external stress may be to promote either boundary displacements or domain vector rotations. For 180° boundary displacements $dl/d\sigma$ is zero. For the other processes it is necessary to know the nature of the internal anisotropy. The simplest case, in which the anisotropy is uniaxial and produced by internal stresses of magnitude σ_i has been considered by Kersten (1933) who finds

$$\Delta\left(\frac{1}{E}\right)=\frac{2}{5}\frac{\lambda_s}{\sigma_i} \qquad \ldots\ldots(9.4)$$

for the rotational process. On the same model the rotational initial susceptibility is $\chi_a=2I_s^2/9\lambda_s\sigma_i$. Thus

$$\Delta\left(\frac{1}{E}\right)=\frac{9}{5}\frac{\lambda_s^2\chi_a}{I_s^2}$$

from which it follows that $\Delta(1/E)\propto\chi_a$ since λ_s and I_s are constants. This point has already been noted as an experimental fact for nickel.

Formulae for $\Delta(1/E)$, both for single and polycrystals have been given by Becker and Döring (1939) when the anisotropy differs from uniaxial. These are all of the form $\Delta(1/E)=a\lambda_s^2/C$ as mentioned previously. (Equation (9.4) takes this form when numerator and denominator are multiplied by λ_s.) Each formula for $\Delta(1/E)$ has its counterpart for χ_a and it is possible to show that for each type of anisotropy and domain process the relation $\Delta(1/E)\propto\chi_a$ holds, i.e. it is possible to eliminate the specific nature of the domain process responsible for the ΔE-effect by means of the known formulae for the initial susceptibility.

The fact that the magnitude of $\Delta(1/E)$ is dependent on the initial domain distribution makes possible the employment of the ΔE-effect to study preferred orientation of domains if λ_s and C are known. This has been done with some success by Williams, Bozorth and Christensen (1941). However the effect appears

to show an even more promising application than this. It has been mentioned that all expressions for $\Delta(1/E)$ are of the same form. The constants a, λ and C will depend on the particular magnetic process and the type of anisotropy governing it. Thus it is possible, since the expressions for $\Delta(1/E)$ can be calculated in all cases, to use measurements of the ΔE-effect to investigate the nature of the predominant magnetic processes in the demagnetized state. A first attempt to interpret the ΔE measurements in this way has been made by Lewis (1954).

The variation of Young's modulus with magnetization below saturation has been treated by Brown (1937), Takagi (1939) and others who show that $\Delta E \propto I^2$. This relation is obeyed by quite a large number of materials. For further details reference should be made to these two papers and the book by Bozorth.

§ 10. Miscellaneous Topics

In this section we group together a number of topics which, either because they are less fundamental or because our knowledge of them is still fragmentary, hardly justify separate treatment.

In the first place we may mention the use of magnetostriction measurements to investigate the domain structure of the demagnetized state. The quantity most often measured is the saturation magnetostriction, i.e. the total strain between the demagnetized and saturated states. Owing to the fact that the demagnetized state is, from the magnetostriction point of view, indeterminate, such measurements do not give reliable information concerning the magnetostriction constants. However if these are known or can be determined by independent investigations then the saturation magnetostriction data can be used to investigate the nature of the domain vector distribution in the demagnetized state (Lee 1955). In general however they are not known and additional supplementary measurements are required. The whole problem has been treated in detail by Rado (1954) who gives the number of observations required for a complete analysis of the degree of preferred domain orientation. The use of magnetostriction measurements to study the effects of magnetic annealing has been made by Williams, Bozorth and Christensen (1941), Hoselitz and McCaig (1949), Shturkin and Shur (1949, 1950) and others. Much of this work has been criticized by Rado (1954).

The problem of the determination of internal strain in cold-worked materials has been studied by a number of methods, using x-ray diffraction techniques. In ferromagnetic substances the presence of internal strains gives rise to a local internal anisotropy of magnitude $\frac{3}{2}\lambda_s\sigma_i \sin^2\phi$ (cf. equation (6.2)) where σ_i is the magnitude of the stress acting at any point inside the material. If this energy is much greater than the crystal energy then magnetic processes will be controlled almost exclusively by the former and formulae may be derived for a number of magnetic properties in terms of λ_s and σ_i. If λ_s is known measurements of these quantities can be used to evaluate σ_i. The quantities most often employed are the initial susceptibility, $\chi_a = \frac{2}{3}I_s^2/\lambda_s\sigma_i$ the change of remanent magnetization with external tension, $dI_R/d\sigma = I_s/4\sigma_i$, and the ΔE-effect, equation (9.4). These are all derived on the basis of the strain theory developed by Becker and Kersten and suffer from all the defects of that theory. On the other hand the measured quantities give estimates which are self consistent and which seem to be in reasonable agreement with estimates obtained by other methods. There are two great weaknesses in the method however. In the first place it cannot be applied to any material

unless it has a very small crystal anisotropy. The second one stems from one of the weaknesses of the strain theory on which the above equations are derived, namely that σ_i does not have the same meaning in all cases and so the actual significance of the value of σ_i obtained is uncertain. This point has been discussed by Köster (1948). For further information reference should be made to the book by Bozorth (1951) and the article by Becker (1940).

No reference has so far been made to the magnetostrictive properties of ferrites and antiferromagnetics. Ferrites possess an essentially cubic structure and so the equations developed in § 2.2 for cubic metals also apply to ferrites. The available data on ferrites is at the moment scanty, being for the most part published either in brief notes or remaining in the relative obscurity of private research reports. The very large moment associated with each lattice site in ferrites will presumably give rise to a dipolar magnetostriction an order of magnitude greater than that existing in ferromagnetics. However recent measurements by Bozorth and Walker (1952) who find $\lambda_{100} = -515 \times 10^{-6}$ for cobalt ferrite, indicate that there must be some other contribution in addition to the dipolar magnetostriction just as in ordinary ferromagnetics. It should also be mentioned that the single crystal elastic constants are not known for ferrites so an exact comparison of the internal forces responsible for magnetostriction in ferrites and ferromagnetics is not yet possible.

The situation is somewhat different for antiferromagnetics. On account of the antiferromagnetic alignment of atomic moments which exists in such substances the net moment of any part of it is zero (unlike ferromagnetics where a sufficiently small element will always be found to be spontaneously magnetized). An external magnetic field has therefore virtually no effect on it and the usual methods of investigating magnetostriction applicable to ferromagnetics are of no use. The existence of a volume magnetostriction was demonstrated by Foex (1948) who showed that the thermal expansion coefficients of MnO, FeO, CoO and NiO exhibited sharp maxima at the Néel point. X-ray investigations by Rooksby (1948) on NiO and by Tombs and Rooksby (1950) on MnO, FeO and CoO have revealed the existence of lattice distortions at temperatures below the Néel point where antiferromagnetic alignment is known to set in. In the first three substances the lattice (face-centred cubic above the Néel point) was distorted into rhombohedral but CoO distorts into a tetragonal lattice. It is natural to regard these distortions as analogous to the linear magnetostriction in ferromagnetics, though the distortion is much greater in antiferromagnetics. These results and others have been discussed in some detail by Lidiard (1954).

Néel (1948 a) has put forward several arguments for the existence of antiferromagnetic domains. These differ from ferromagnetic domains in that their net moment is zero so they are not oriented by a magnetic field. Bozorth (1951) has suggested that movements of domain walls might be brought about by external stresses. Street and Lewis (1951) and Fine (1952) have observed large changes in Young's modulus at the Néel point of CoO and NiO. Fine has interpreted this as evidence for the existence of a domain structure and for stress induced domain wall movements, but in view of the very considerable departures from cubic symmetry below the Néel point it is by no means certain that this is the only explanation.

§ 11. Conclusion

In this report an attempt has been made to present an account of the basic phenomena underlying the various observed magnetomechanical effects, and to

give, as far as possible, a coherent picture of the present state of the subject. In doing this it was necessary to bear in mind the fact that no other report on magnetostriction, of any magnitude, has appeared since McKeehan's article of 1926. On the other hand the subject is treated fairly fully in the often quoted books by Becker and Döring where the emphasis is rather on the theoretical side, and by Bozorth which deals very largely with the experimental aspects. Since the latter book appeared only quite recently, a good deal of material upon which emphasis was therein laid has been purposely omitted here. This includes, chiefly, the detailed behaviour of ferromagnetics under stress in the presence of applied magnetic fields, and what Bozorth terms ' magnetostriction data '. For a similar reason no mention has been made of magnetomechanical effects at any other than zero frequency. A very complete discussion of dynamic magnetostriction and of systems undergoing magnetostrictive oscillations has been given by Van der Burgt (1953) whose paper contains an extensive bibliography.

No mention has been made of experimental procedure except in cases where this is particularly difficult. In the measurement of the magnetostriction constants this is not so, and the best technique, using electrical resistance strain gauges, is adequately described by Goldman and Smoluchowski (1949). The best procedure for measuring the magnetostriction constants when single crystal specimens are available is described in detail by Bozorth and Hamming (1953).

There is an enormous literature on the subject of magnetostriction and the reference list included below aims to be selective rather than comprehensive. Order was brought into the existing chaos by the appearance of the papers by Akulov (1928) and Becker (1930), and it may safely be said that with the exception of the papers by McKeehan and his school, and certain of the Japanese workers, little of the literature written before 1928 is of any great value. With regard to the work published after this date, although the reference list does not pretend to be complete it is nevertheless hoped that there are no serious omissions.

Acknowledgments

The author would like to express his thanks to Professor L. F. Bates for his advice in the writing of this report. He is indebted to Drs. J. E. Goldman and B. Lewis for sending him accounts of their work in advance of publication, and to all those who sent reprints of their published work.

References

Alizade, Z. I., 1950, *C.R. Acad. Sci., U.R.S.S.*, **73,** 79.

Akulov, N. S., 1928, *Z. Phys.*, **52,** 389; 1930, *Ibid.*, **66,** 533; 1931, *Ibid.*, **69,** 78; 1939, *Ferromagnetism* (Moscow, Leningrad) (in Russian).

Von Auwers, O., 1933, *Phys. Z.*, **34,** 824; 1944, *Ibid.*, **45,** 192.

Averbach, B. L., and Warren, B. E., 1952, *Imperfections in nearly Perfect Crystals* (London: Chapman and Hall), p 152.

Azumi, K., and Goldman, J. E., 1954, *Phys. Rev.*, **93,** 630.

Becker, R., 1930, *Z. Phys.*, **62,** 253; 1934, *Ibid.*, **87,** 547; 1940, *Proc. Phys. Soc.*, **52,** 138.

Becker, R., and Döring, W., 1939, *Ferromagnetismus* (Berlin: Springer) (Photo-lithoprint reproduction, 1943, Ann Arbor, Michigan: Edwards Brothers).

Becker, R., and Kersten, M., 1930, *Z. Phys.*, **64,** 660.

Becker, R., and Kornetzki, M., 1934, *Z. Phys.*, **88,** 634.

Bethe, H., 1933, *Handbuch der Physik* (Berlin: Springer), Vol. 24.

Bitter, F., 1937, *Introduction to Ferromagnetism* (New York: McGraw-Hill).

BLOCH, F., and GENTILE, G., 1931, *Z. Phys.*, **70,** 395.
BOZORTH, R. M., 1951, *Ferromagnetism* (New York: Van Nostrand); 1953, *Rev. Mod. Phys.*, **25,** 42; 1954, *Phys. Rev.*, **96,** 311.
BOZORTH, R. M., and HAMMING, R. W., 1953, *Phys. Rev.*, **89,** 865.
BOZORTH, R. M., MASON, W. P., and MCSKIMIN, H. J., 1951, *Bell Syst. Tech. J.*, **30,** 970.
BOZORTH, R. M., and SHERWOOD, R. C., 1954, *Phys. Rev.*, **94,** 1439.
BOZORTH, R. M., and WALKER, J. G., 1952, *Phys. Rev.*, **88,** 1209.
BRAUNEWELL, W. and VOGT, E., 1949, *Z. Naturf.*, **4A,** 491.
BROCKHOUSE, B. N., 1953, *Canad. J. Phys.*, **31,** 339.
BROOKS, H., 1940, *Phys. Rev.*, **58,** 909.
BROWN, W. F., 1937, *Phys. Rev.*, **52,** 325; 1953, *Rev. Mod. Phys.*, **25,** 131.
CARR, W. J., and SMOLUCHOWSKI, R., 1951, *Phys. Rev.*, **83,** 1236.
CHEVENARD, P., 1927, *Trav. Bur. Int. Poids Mes.*, **17.**
DÖRING, W., 1936, *Z. Phys.*, **103,** 560; 1938, *Ann. Phys., Lpz.*, **32,** 465.
DYAKOV, G. P., 1947, *Bull. Acad. Sci., U.R.S.S.* (*Phys.*), **11,** 667; 1951, *C.R. Acad. Sci. U.R.S.S.*, **76,** 201.
EBERT, H., and KUSSMANN, A., 1937, *Phys. Z.*, **38,** 437.
ELSAS, A., and VOGT, E., 1951, *Z. Naturf.*, **6A,** 223.
ENGLER, O., 1938, *Ann. Phys., Lpz.*, **31,** 145.
FINE, M. E., 1952, *Phys. Rev.*, **87,** 1143; 1953, *Rev. Mod. Phys.*, **25,** 158.
FOEX, M., 1948, *C.R. Acad. Sci., Paris,* **227,** 193.
FOWLER, R. H., and KAPITZA, P., 1929, *Proc. Roy. Soc.* A, **124,** 1.
FÖRSTER, F., and KÖSTER, W., 1937, *Naturwissenschaften*, **25,** 436.
GANS, R., 1932, *Ann. Phys., Lpz.*, **15,** 28.
GANS, R., and VON HARLEM, J., 1933, *Ann. Phys., Lpz.*, **16,** 162.
GOLDMAN, J. E., 1947, *Phys. Rev.*, **72,** 529; 1949, *Ibid.*, **76,** 471; 1950, *Ibid.*, **80,** 301
GOLDMAN, J. E., and SMOLUCHOWSKI, R., 1949, *Phys. Rev.*, **75,** 140.
GUGGENHEIM, E. A., 1936, *Proc. Roy. Soc.* A, **155,** 49.
GUILLEMIN, A., 1846, *C.R. Acad. Sci., Paris,* **22,** 264, 432.
HEISENBERG, W., 1931, *Z. Phys.*, **69,** 287.
HIDNERT, P., 1930, *Bur. Stand. J. Res., Wash.*, **5,** 1305.
HOLSTEIN, T., and PRIMAKOFF, H., 1940, *Phys. Rev.* **58,** 1098.
HONDA, K., and TERADA, T., 1907, *Phil. Mag.*, **13,** 36; *Ibid.*, **14,** 65.
HOSELITZ, K., and MCCAIG, M., 1949, *Proc. Phys. Soc.* B, **62,** 163.
JONES, G. O., and STACEY, F. D., 1953, *Proc. Phys. Soc.* B, **66,** 255.
JOULE, J. P., 1842, *Sturgeons Annals of Electricity*, **8,** 219.
KAPITZA, P., 1931, *Proc. Roy. Soc.* A, **131,** 224.
KATAYAMA, T., 1951, *Sci. Rep. Res. Insts Tôhoku Univ., Ser.* A, **3,** 341.
KAYA, S., and TAKAKI, H., 1935, *J. Fac. Sci. Hokkaido Univ.*, **1,** 227.
KERSTEN, M., 1933, *Z. Phys.*, **85,** 708.
KIMURA, R., and OHNO, K., 1934, *Sci. Rep. Tôhoku Univ.*, **23,** 359.
KIRCHNER, H., 1936, *Ann. Phys., Lpz.*, **27,** 49.
KIRKHAM, D., 1937, *Phys. Rev.*, **52,** 1162.
KITTEL, C., 1949, *Rev. Mod. Phys.*, **21,** 541.
KORNETZKI, M., 1934, *Z. Phys.*, **87,** 560; 1935, *Ibid.*, **98,** 289; 1938, *Wiss. Veröff. Siemens*, **17,** 410.
KÖSTER, W., 1943, *Z. Metallk.*, **35,** 194; 1948, *Z. Phys.*, **124,** 545.
KÖSTER, W., and SCHMIDT, W., 1934, *Arch. Eisenhüttenw.*, **8, 25.**
KUSSMANN, A., and GR. VON RITTBERG, 1950, *Ann. Phys., Lpz.*, **7,** 173.
LAWTON, H., and STEWART, K. H., 1948, *Proc. Roy. Soc.* A, **193,** 72; 1950, *Proc. Phys. Soc.* A, **63,** 838.
LEE, E. W., 1954, *Proc. Phys. Soc.* A, **67,** 381; 1955, *Ibid.*, **68,** 65.
LEWIS, B., 1954, *Thesis*, Nottingham University.
LIDIARD, A. B., 1954, *Rep. Prog. Phys.*, **17,** 201 (London: Physical Society).
LOVE, A. E. H., 1936, *Mathematical Theory of Elasticity*, 4th edn. (Cambridge: University Press).
MASIYAMA, Y., 1928, *Sci. Rep. Tôhoku Univ.*, **17,** 945.
MASON, W. P., 1950, *Phys. Rev.*, **82,** 715; 1954, *Ibid.*, **96,** 302.

MASON, W. P., and LEWIS, J. A., 1954, *Phys. Rev.*, **94,** 1439.
MCCAIG, M., 1952, *Nature, Lond.*, **169,** 889.
MCKEEHAN, L. W., 1926, *J. Franklin Inst.*, **202,** 737.
MESSKIN, V. S., SOMIN, B. E., and NEKHAMKIN, A. S., 1941, *J. Tech. Phys., Moscow*, **11,** 918.
MOTT, N. F., and JONES, H., 1936, *Theory of the Properties of Metals and Alloys* (Oxford: Clarendon Press).
MUELLER, H., 1940, *Phys. Rev.*, **58,** 805.
NÉEL, L., 1937, *Ann. Phys., Paris,* **8,** 237; 1939, *Le Magnétisme* (Strasbourg: C.N.R.S.) Vol. 2, p. 65; 1944, *J. Phys. Radium*, **5,** 241; 1948 a, *Ann. Phys., Paris*, **3,** 137; 1948 b, *J. Phys. Radium*, **9,** 193; 1954 a, *Ibid.*, **15,** 225; 1954 b, *Ibid.*, **15,** 376.
PATRICK, L., 1954, *Phys. Rev.*, **93,** 384.
PIGEON, H. A., 1919, *Phys. Rev.*, **13,** 209.
POLLEY, H., 1939, *Ann. Phys., Lpz.*, **36,** 625.
POWELL, F. C., 1931, *Proc. Camb. Phil. Soc.*, **27,** 561.
RADO, G. T., 1954, *J. Appl. Phys.*, **25,** 102.
ROOKSBY, H. P., 1948, *Acta Cryst., Camb.*, **1,** 226.
SCHARFF, G., 1935, *Z. Phys.*, **97,** 73.
SEITZ, F., 1934, *Z. Kristallogr.*, **88,** 433.
SHTURKIN, D. A., 1947, *Bull. Acad. Sci., U.R.S.S.*, **9,** 661.
SHTURKIN, D. A., and SHUR, J. S., 1949, *J. Tech. Phys., Moscow*, **19,** 599; 1950, *Ibid.*, **20,** 1393.
SIEGEL, S., and QUIMBY, S. L., 1936, *Phys. Rev.*, **49,** 663.
SLATER, J. C., 1930, *Phys. Rev.*, **36,** 57.
SMOLUCHOWSKI, R., 1941 a, *Phys. Rev.*, **59,** 307; 1941 b, *Ibid.*, **60,** 249.
SNOEK, J. L., 1937, *Physica*, **4,** 853; 1947, *New Developments in Ferromagnetic Materials* (Amsterdam: Elsevier).
STEWART, K. H., 1949, *Proc. Camb. Phil. Soc.*, **45,** 296.
STONER, E. C., 1935, *Phil. Mag.*, **19,** 565; 1937, *Ibid.*, **23,** 833; 1948, *Rep. Prog. Phys.*, **11,** 43; 1950, *Ibid.*, **13,** 83.
STREET, R., and LEWIS, B., 1951, *Nature, Lond.*, **168,** 1036.
SUCKSMITH, W., 1950, *B.E.A.I.R.A., Report N/T51.*
TAKAGI, M., 1939, *Sci. Rep. Tôhoku Univ.*, **28,** 20.
TAKAKI, H., 1937, *Z. Phys.*, **105,** 92.
TOMBS, N. C., and ROOKSBY, H. P., 1950, *Nature, Lond.*, **165,** 442.
VAN DER BURGT, C. M., 1953, *Philips Res. Rep.*, **8,** 91.
VAN VLECK, J. H., 1937, *Phys. Rev.*, **52,** 1178.
VILLARI, E., 1865, *Ann. Phys., Lpz.*, **126,** 87.
VLADIMIRSKY, K. V., 1943, *C.R. Acad. Sci., U.R.S.S.*, **41,** 10.
VONSOVSKY, S. V., 1940, *J. Phys., Moscow*, **3,** 181.
WEISS, P., 1907, *J. Phys. Théor. Appl.*, **6,** 661.
WENT, J. J., 1951, *Physica, Eindhoven*, **17,** 98.
WIEDEMANN, W., 1883, *Lehre von der Elektrizität* (Braunschweig: Vieweg) Vol. 3, 680.
WILLIAMS, C., 1934, *Phys. Rev.*, **46,** 1011.
WILLIAMS, H. J., BOZORTH, R. M., and CHRISTENSEN, H., 1941, *Phys. Rev.*, **59,** 1005.
WILLIAMS, S. R., 1931, *Magnetic Phenomena* (New York: McGraw-Hill).
YAMAMOTO, M., and MIYASAWA, R., 1953, *Sci. Rep. Res. Insts Tôhoku Univ.* A, **5,** 113.

ELECTROSTRICTION

By H. F. KAY

H. H. Wills Physical Laboratory, University of Bristol

CONTENTS

		PAGE
§ 1.	Introduction	230
§ 2.	General properties of the materials	231
§ 3.	Piezoelectricity and electrostriction	234
§ 4.	Growth of single crystals	239
§ 5.	Striction in single crystals	241
§ 6.	Striction in ceramic materials	243
	References	249

Abstract. The term ferroelectricity is briefly introduced and the properties of the perovskite-structure mixed-oxide group are reviewed with special reference to $BaTiO_3$. The distinction between piezoelectricity and electrostriction is explained in some detail, and the relevant expressions relating the electrical and mechanical properties are discussed for both the single crystal and ceramic specimens. It is pointed out that there is still considerable doubt as to the exact mechanism involved, especially in the case of the ceramics when electrical stresses are applied, and simplified models are discussed to demonstrate the factors contributing to the resultant strain.

Single crystals are then discussed in detail, with a survey of the methods of growth and results of striction and hysteresis loop measurements; the most likely interpretation of these results is discussed. Similarly the more complicated ceramic materials are dealt with, and it is shown that measured values of the striction constants conform reasonably with theory, but further analysis is still needed to account for small discrepancies. Time dependent effects are also described and tentative explanations given.

Brief mention is made of other oxide materials and possible directions of future research. (56 references).

§ 1. Introduction

With the development of the new perovskite type ceramic materials, over the last decade, interest in the whole subject of ferroelectrics has received additional impetus, and a considerable range of technically important electrical materials has resulted.

It was natural that the first property of these new ceramics to receive attention, as in the case of other oxides like TiO_2, was their extremely high dielectric constant, but it soon became apparent that their electrostrictive properties showed promise in fields hitherto reserved almost exclusively for single crystals of piezoelectric materials such as quartz, or piezoelectric domain substances such as the phosphates isomorphous with potassium dihydrogen phosphate (KH_2PO_4) and tartrates isomorphous with rochelle salt ($NaKC_4H_4O_6.4H_2O$).

Only in the case of piezoelectric materials involving no domain orientation

effects can the strictive properties be discussed in a simplified manner somewhat divorced from other electrical characteristics, for in these materials the relationship between electric field, polarization, mechanical stress and strain can be considered constant over the range of most practical application.

This is no longer the case in ferroelectrics and an understanding of the other non-linear characteristics such as hysteresis phenomena and field dependence effects is needed before a reasonable understanding of their properties can be developed.

The term striction in the electric rather than the magnetic case usually refers to a strain effect which is proportional to the square (and even orders) of the applied field. Although present in all materials, this effect by itself is usually extremely small, and has little practical application. The strain effect which is directly proportional to the applied electric field, if present at all, is often many orders larger and may be of sufficient magnitude for use in electromechanical transducers. This is known as the piezoelectric effect and has no exact magnetic analogue, although the closest approach to this would be the striction observed in a saturated permanent magnet. (Such is observable with high biasing magnetic fields, but is very small compared with the ordinary magnetic strain effect of value of the order of 10^{-5} associated with domain reversal.)

Quartz is the best known example of the truly piezoelectric materials and such substances will not be considered in this discussion. Their properties and applications are well established and are dealt with fully in such extended works as those of Cady (1946) and Mason (1951 a), while an abundance of texts exist concerning the applications of quartz, written by Vigoureux (1939, 1950) and others. Progress in this field lies in cutting and lapping methods, detwinning, the process of ageing, and the growth of synthetic crystals by hydrothermal processes (Thomas *et al.* 1949, Brown *et al.* 1952).

It is those materials having large piezoelectric coefficients by virtue of their reversible electrical domains, or with nonpolar–polar structural instability which are to be considered here. This is the group of materials known as ferroelectrics, and the ceramic oxide group offers perhaps the best opportunities to date of obtaining high strictive coefficients, together with greater ease of fabrication, compared with true piezoelectric materials having irreversible polarity.

Ferroelectric materials themselves can again be divided into several groups delineated essentially by their structures and to some extent by the present beliefs of the mechanism of their behaviour. The perovskite–ilmenite group, rochelle salt group and the dihydrogen phosphate group are the divisions most commonly employed, although when the structures are considered in detail further subdivisions become apparent, such as ferroelectric and anti-ferroelectric, or classification by crystal symmetry. The emphasis will be placed on the first group because of its very striking advantages in such properties as resistance to water, chemical action and thermal shock, greater hardness, and above all its effectiveness for powder compresses which can be preformed by ceramic powder sintering techniques, thus eliminating the expensive processes of crystal growing and cutting.

§ 2. General Properties of the Materials

Barium titanate ($BaTiO_3$) is the material which has been studied in greatest detail, but the structure-type is to be found in many double oxides of the type ABO_3

where A is (Pb.Ba.Sr.Ca.Cd) or (Li.Na.K) and B is (Ti.Zr.Sn) or (Nb.Ta). Other elements, such as Mg and Zn have been added in small proportions to alter their properties. Not all the compositions outlined show ferroelectric properties of magnitude sufficient for practical applications, and in fact many of the possible mixture materials (i.e. those containing more than one type of A or B ion) have not been made at all. However, some progress is being made in the study of the group as a whole to obtain insight into the reasons for their individual characteristics. The mixtures approaching the composition $BaTiO_3$ have greatest use at present, the small admixtures being made to reduce variation of properties with temperature and time, thus allowing the production on a commercial scale of stabilized electrical components such as decoupling condensers, transducer elements, dielectric amplifiers, memory devices etc.

The general properties of ceramic ferroelectric materials based on $BaTiO_3$ have been discussed frequently, and useful summaries are to be found in the literature (Jackson 1950, von Hippel 1950, Devonshire 1954). It is intended here only to remind the reader of the salient features and properties in order to facilitate the understanding of the information presented.

These materials have high dielectric constant, frequently reaching a peak value at or near a so-called Curie temperature where a first- or second-order phase change occurs in which no structural breakdown is apparent (i.e. external morphology is maintained). This change may result in a parallel or antiparallel array of dipoles, which are either statistically random or even non-existent above the Curie temperature, and which may be seriously affected below the transition temperature by practically applicable electric fields or mechanical stress. The symmetry change involved at the transition invariably produces changes in strictive constants. These constants are unusually high, due to the inherent symmetry instability associated with the relative movement of ions essential to the occurrence of sharp maxima in the variation of dielectric constant with temperature.

The wide variety of mechanisms outlined above, which may possibly occur in a large range of chemical compositions having essentially similar structure, all combine to give a range of properties as yet hardly interpreted in anything but coarse detail.

$BaTiO_3$ has large values of piezoelectric moduli disposed within a crystal symmetry 4 mm which is relatively easy to study in that only four constants are involved. This property, together with the often confused but similar mechanism of electrostriction, is already being put into practice, but much improvement is still necessary in the techniques of pre-polarizing or ' poling ', i.e. the process of inducing as much as possible of the single crystal properties into the easily produced ceramic material by suitably directed electric fields.

Resistance to water absorption is similar to that of any other ceramic material, but an important practical difficulty is the production of sufficient density in the materials without affecting chemical composition detrimentally to the loss factor in particular. Careful work is required in this respect by way of particle grading and the mechanics of particle flow during pressing.

The disadvantages of these materials, some of which may occasionally be put to practical use, include the extreme temperature dependence of all these properties: dielectric constant, loss, piezo and strictive effects, with the additional difficulty over the magnetostrictive materials resulting from the very large anisotropy (10^{-2} or

more compared with the value of 10^{-5} for magnetic materials) which severely limits a parallel interpretation.

Perhaps one of the most serious disadvantages of these materials is that they do not lend themselves readily to the production of single crystals, due chiefly to their refractory nature and chemical inertness at lower temperatures. As a result, much of the research on these materials has been impeded by the anomalous effects of polycrystallinity, e.g. the almost inevitable composition changes across the small grains and resultant changes in elastic and electrical properties, which have received little attention up to the present—not that it is suggested that the complete analysis of the properties of the single crystal would present an immediate and easy interpretation of the polycrystalline mass (a start has been made in this direction by Devonshire (1949, 1951)) but this problem is considerably easier than the converse. The interpretation of the bulk properties must of necessity involve a full understanding of the single crystal if undue empiricism is to be avoided.

Apart from the advantages related to the elucidation of the basic properties and mechanisms of these materials, single crystals will probably have direct practical application by virtue of their larger piezo coefficient (of the order of ten times better than the usual ceramic) and their more rectangular hysteresis loop which, for instance, is essential for memory devices. In this connection it might be mentioned that on crystals 62 mm^2 area and 0·2 mm thick, an information density of 4 'bits' per mm^2 has been achieved using 0·5 μsec recording and read-out pulses, and a 'dit–dot' discrimination of 10 to 1 in energy, the memory (polarization) lasting several days (Anderson 1952). However it is still doubtful whether ferroelectrics or 'square loop' ferrites will eventually prove more practicable in this field.

Many of the properties of these materials find their parallel in the realm of ferromagnetics, for example, the existence of domains, their formation due to co-operative phenomena at a Curie temperature and their changes in dimensions due to the application of fields; but outstanding differences are apparent which make close parallels dangerous.

In the magnetic case, free poles do not exist, so that permeability effects are concerned essentially with domain wall motion and eddy currents, whereas the ferroelectrics differ by the possible existence of free poles at crystallite boundaries and the compensation of external fields by free charges.

The ordering force in the electrical case results from the summation over several interpenetrating lattices of long-range interatomic effects which gives a local field contribution greater by a factor of from 4 to 20 over the usually accepted value of $4\pi P/3$ as used in the simple Clausius–Mosotti relation for ordinary dielectrics (Slater 1950). In ferromagnetics the ordering force, although not so clearly understood, is believed to be essentially short range in nature between adjacent atoms, and results from the energy reduction effected by spatially sensitive electron orbitals. Only in the case of a few face-centred salts of the type used in adiabatic demagnetization (e.g. some mixed alums) may any serious comparison be made between the origin of the local field in the ferroelectric and ferromagnetic cases (Luttinger and Tisza 1946, Sauer 1940). In general then, the ordering forces in the ferroelectrics are not expected to be as high as in the ferromagnetics, which accounts in some measure for the lower Curie temperatures, the higher incremental permittivity at saturation and the lower Curie constant of ferroelectric materials.

In considering the long-range co-operative effect occurring in the ferroelectrics,

one should not overlook the various possible short-range interactions which might include simple ionic packing considerations (Kay 1947), and the tendencies for separate oxide groupings due to polar bond formation (Eucken and Büchner 1934, von Hippel 1952, Megaw 1952, 1954). These ideas do not seriously impair the theoretical arguments of Slater and Devonshire concerning the behaviour of the interaction, irrespective of the exact nature of the latter.

§ 3. Piezoelectricity and Electrostriction

The application of an electric field to any perfectly centro-symmetrical arrangement of atoms, ions or complexes (e.g. NaCl) will move these into non-symmetrical positions and result in a homogeneous distortion, more usually a contraction in the direction of the field together with an expansion at right angles through what might loosely be called Poisson coupling. Reversal of the field must, because of symmetry, produce the same sign of striction, so that the strain can be written in the form $x=bE^2+dE^4$ + higher even powers. Usually the first term is sufficient in practice, the higher terms being negligible for applied fields within the breakdown strength, but obviously higher order anharmonic terms must be present in the coupling of the constituent ions which, among other properties, accounts for that of thermal expansion. In the case of cubic $BaTiO_3$, for example, the departure of the oxygen potential wells from being truly harmonic is one of the important features of the theory (Devonshire 1949, Slater 1950) and is used to explain the observed temperature dependence of the dielectric susceptibility, so that the higher order terms in the striction equation may be more significant in this case. However, one must beware of confusing the striction obtained with cubic materials and the effect obtained with a ceramic possessing rotatable domains, for the latter usually produces a strain dependent on even powers of field and this may bear little relation to the atomic force constants of the single domain except in so far as the anisotropy restricts the domain orientation.

If the production of strain by the application of a field is known as the ' inverse ' effect, then the ' direct ' effect, the production of a resultant polarization by the application of a stress is impossible in centro-symmetrical materials since any homogeneous mechanical distortion of a centro-symmetrical array of vectors (e.g. dipoles) does not destroy the centre of symmetry. The distinction between applied field and stress results from the one acting on charge points of which both +ve and −ve exist, while the other acts only on mass points.

The application of a biasing field to an electrostrictive material will produce an asymmetrical atomic arrangement, and this in turn will result in an asymmetric distortion by subsequent fields which are small compared with the bias. This is an incremental striction which reverses with the incremental field E and is of the form $x=aE+cE^3+$higher odd powers. Provided the incremental field is much less than the biasing field only the linear term is significant.

The origin of this biasing field may be of three kinds in a truly single crystal: firstly, the interatomic co-operative ordering forces as occur in single domains of ferromagnetic and ferroelectric materials, secondly, the much stronger effect of steric hindrance which obtains in truly irreversible polar crystals whose reversal would require the complete breakdown of the structural bonding. Examples of the second case are boracite and tourmaline which are permanent electrical dipoles and have no magnetic analogue since the magnetic vector cannot be jammed by steric

hindrance nor can a molecule have strongly magnetically polar bonds. Finally the biasing field may be applied externally, but the resulting strain is usually too small to have much significance except when unstable materials like ferroelectrics are biased just above their transition temperature. In cases where domains can be reoriented by bias, a degree of asymmetry is produced depending on the impediments to domain reversal, such as crystalline defects, charge relaxation, macrostrains and volume limitations. Again the distinction between a single centro-symmetrical crystal and an aggregate of asymmetric polycrystals should be clearly borne in mind.

These naturally biased materials then are piezoelectric and the direction of the resultant polarization will depend on the crystal symmetry and direction of the applied stress according to the relations

$$P_m = -d_{mh}X_h + \eta_{mk}{}^X E_k \qquad \text{(direct effect)}$$

$$x_h = d_{hm}E_m - S_{hi}{}^E X_i \qquad \text{(indirect effect)}$$

where either of the subscripts k and m can have values 1, 2, 3, which represent the three principal components of polarization P or field E, and either of the subscripts h and i, the six components of a symmetrical second order stress X or strain x tensor; η^X is the dielectric susceptibility at constant stress and S^E the elastic compliance at constant field. In practice the simpler type of experiment would be sought so that the last term in each of the above equations could be neglected. The piezoelectric coefficient $d_{mh}(=d_{hm})$ has 18 components in the general case; these reduce in number as the crystal symmetry increases. Such expressions relate entirely and simply to those materials exhibiting piezoelectric effects clearly distinct from electrostriction. Quartz is an obvious example, but we may also include those ferroelectric materials like rochelle salt and the dihydrogen phosphates above their transition temperatures where a large piezoelectric coefficient still exists, for although their d values change anomalously as the transition temperature is passed, these anomalies are understandable and easily calculable after a model has been chosen to explain the gradual change of clamped dielectric susceptibility $(P/E)_x$ with temperature which reacts through a temperature-independent ratio E/X to give a rapid variation of piezoelectric constant P/X. In other words, a model to account for one anomaly will explain the others by normal electromechanical coupling equations true for all types of crystals (see Devonshire 1954, p. 97).

In the case of $BaTiO_3$ class of materials, however, the situation is somewhat different in that the symmetry is cubic above the transition temperature and therefore this material can exhibit no dependence of strain on odd powers of E or P, whereas below the transition it is tetragonal and can have piezoelectric constants d_{33}, $d_{23}=d_{13}$, $d_{51}=d_{42}$ where axis 3 corresponds to the axis of spontaneous polarization (crystal c axis). Thus as the temperature is raised these d coefficients will vary rapidly as the anisotropy (spontaneous polarization) varies, and they must eventually disappear, to be replaced by striction coefficients which will appear to the observer as anomalous piezoelectric effects, the continuity of the curve being maintained by the influence of the applied field on the value of the transition temperature itself.

It is preferable then in the case of a single domain of $BaTiO_3$ type materials to refer to all the effects as a striction of the cubic $BaTiO_3$ represented by an equation of the type $x_h=Q_{hmn}P_mP_n$ (or more accurately $x_h=-S_{hi}X_i+Q_{hmn}P_mP_n$ together with

an equivalent stress equation for X_h) where P_m and P_n are to include both the spontaneous polarization P_{0m} and the external field-induced polarization P_m etc. At temperatures ranging from $-120°$C where the spontaneous polarization is along the z or 3 axis, this expression can be written

$$x_h = Q_{h3n}(P_{03}+P_3)\,(P_{0n}+P_n)$$

and the only relevant terms for paraxial fields are

$$x_3 = Q_{333}(P_{03}+P_3)\,(P_{03}+P_3) \text{ for field applied along } 3$$

$$x_{\underset{1}{2}} = Q_{\underset{133}{233}}(P_{03}+P_3)\,(P_{03}+P_3) \text{ " " " " } 3$$

$$x_{\underset{5}{4}} = Q_{\underset{532}{431}}(P_{03}+P_3)P_{\underset{2}{1}} \text{ " " " " } 2$$

and we can write $Q_{333}=Q_{33}$, $Q_{233}=Q_{23}$ and $Q_{431}=Q_{43}$. This produces zero shear strain above the transition since $P_{03}+P_3=0$. The piezoelectric nature of the effect below the Curie temperature is apparent if we consider $P_{03} \gg (P_3$ applied) such that a piezoelectric constant of the type QP_{03} exists, given by $x_3=(2Q_{33}P_{03})P_3$ etc.

So far the discussion has related to the constants of a single domain. Direct measurement of these fundamental constants at temperatures below T_c are difficult because of the occurrence of twins (i.e. domains) which would tend to give a mixture of at least two of these coefficients, and also because of the difficulty of deriving the value of P_0 directly from the measurements. It is only possible to measure the Q value for a single crystal over a short range of polarization about a mean equal to the value of the spontaneous polarization. Attempts to encourage greater changes in P will change the domain orientation.

One point at least can be found on the (x, P) curve by utilizing the room temperature strain $x_3=0{\cdot}007$ and $x_1=x_2=0{\cdot}0034$ as obtained from x-rays (Kay and Vousden 1949) and the spontaneous polarization $P_{03}=25\times10^{-6}$ coulombs cm^{-2} (Merz 1953) which gives $Q_{33}=1{\cdot}26\times10^{-12}$ e.s.u. and $Q_{23}=-0{\cdot}66\times10^{-12}$ e.s.u. This latter assumes the volume striction is zero; that a slight volume striction is expected can be seen by the very small volume expansion on cooling through the Curie temperature at 120°C (Kay 1947). It would seem to be more reasonable to calculate the strain from the volume that the unit cell would have had at room temperature as obtained by extrapolating the strain–temperature line which obtains above the Curie temperature. This gives $Q_{33}=1{\cdot}38\times10^{-12}$ e.s.u. and $Q_{23}=-0{\cdot}49\times10^{-12}$ e.s.u. The fact that the strain appears to be roughly proportional to P_0^2 for all temperatures below the upper Curie point as checked by both x-ray and optical birefringence strain measurements (Kay and Vousden 1949) indicates that it may be reasonable to assume that this relation $x \propto P_0^2$ may hold at constant temperature if x and P could be continuously varied and measured. However, this is an assumption which may eventually prove incorrect, for the results to date suffer from the fact that x and P have not been measured on the same crystal.

The next step is to transfer these properties to single crystals containing domains, and finally to the ceramic; it is here that the difficulties arise, although the available measurements are rather more numerous, particularly on the technically important ceramic materials.

Assuming the simplest model in which the crystallite array is ignored, and that the whole block of ceramic is considered as an isotropic unstrained medium simulat-

ing a single crystal, then the (x, P) curve would lie on that shown in figure 1(*a*) giving an observed Q'_{33} equal to the single crystal value $Q_{33}=x_0/P_0^2=1{\cdot}38\times 10^{-12}$ e.s.u. In all probability the curve would not extend as far as the point x_0, P_0 because full saturation is not achieved due to high crystal anisotropy and the clamping of crystallites by neighbours. Any departure from an accurate single crystal P^2 law would be indicated by some other curve such as that shown dotted.

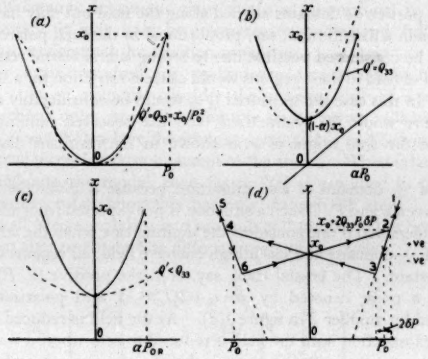


Figure 1

Another simple model consisting of unstrained domains turning completely freely gives a very similar result. A domain polarized at an angle θ to the applied field gives $P_3=P_{03}\cos\theta$ and length of the form $l_3=a+(c-a)\cos^2\theta$. The average value of l_3 for all orientations is $a+(c-a)\overline{\cos^2\theta}=a+(c-a)/3$ so that the strain is

$$x_3=\frac{\cos^2\theta-\frac{1}{3}}{a/(c-a)+\frac{1}{3}}$$

from the randomized position. Inserting values of c and a for the unit cell we find $x_3=(P_3^2/100P_0^2)-0{\cdot}003$ which represents an observed Q'_{33} of $1{\cdot}9\times 10^{-12}$ e.s.u. with a slight displacement of the point of origin which can never be detected anyway. These two simple models suggest that observed values of Q'_{33} etc. should not be much larger than the single crystal values, but other mechanisms must be invoked to account for the different types of curves actually observed.

Polarized ceramics sometimes show asymmetric (x, D) curves. If the model assumes that the rotating elements are diluted by a factor $1-a$ of domains oriented parallel to the field direction by a pre-polarizing field and locked in position by internal forces, then the general shape of the (x, P) curve would conform with the usual shape (but would not extend further than a polarization aP_0) and would be displaced from the origin. Superimposed on this would be the ordinary linear piezoelectric effect of the locked domains giving a straight line having slope$\sim 2(1-a)Q_{33}P_0$. These effects shown in figure 1(*b*) are added giving the resultant dotted curve which is to be expected.

The main feature is that the measured Q'_{33} does not differ seriously from that obtained in the unpolarized case and experiments to date have shown this type of curve, although use of higher a.c. fields to decrease a to zero can eliminate the asymmetry. Hysteresis effects are expected only for (D, E) or (x, E) curves, and these in turn must be some indication of that proportion of rotating component contributing at any given moment.

Another model might assume a proportion of ordinary rotating P_{0R} groups with the remaining portion as domains aligned along the field but only having +ve and −ve directions in a distribution, say, proportional to the total polarization. Such regions might be considered possible due to strong lateral elastic restraint. Then the application of field to these regions would cause no striction for a large change of polarization. In this case the measured Q'_{33} would be considerably reduced while the (x, P) curve would be symmetrical, and the observed minimum would be displaced from the true origin of x as shown in the resultant dashed curve of figure 1(c).

A model of $\pm$ domains is also important because it gives the possibility of hysteresis between x and P. Such a situation is not expected from simpler models. This can be understood if one considers the limiting case when the whole specimen is composed of $\pm$ domains and has a high coercive field, all regions changing sign at the same instant. The crystal state, say, is in the positive (x, P) quadrant of figure 1(d) at a point denoted by $x=x_0+2Q_{33}P_0\delta P$ and polarization of value $P_0+\delta P$ denoted by number 2 in figure 1(d). As the field is reduced the specimen experiences piezostriction until the field is reduced to zero where $x=x_0$ and $P=P_0$. The field then becomes negative and approaches the coercive value when the crystal state jumps from the point 3 where $x=x_0-2Q_{33}P_0\delta P$; $P=P_0-\delta P$ to a point where $P=-P_0-\delta P$; $x=x_0+2Q_{33}P_0\delta P$ and so on. If all the regions do not change over at the same value of the total polarization, the curve is smoothed out, and if ordinary striction is superimposed on this, there results a parabola showing some hysteresis. Such a model may be particularly relevant to the case of ceramics of antiferroelectric materials, or materials just above the Curie temperature, and of materials exhibiting conductivity.

These simple models illustrate the difficulties associated with the interpretation of the striction curves, and progress to date has not reached that certainty achieved for the magnetic materials. All the effects used in the models have been observed but the detailed analysis of the various contributions of rotation of polarization has not been carried out for any individual sample.

In our simple discussion, the contribution due to cross terms Q_{43} etc. have been ignored, together with the important fact that many domains have not their stress-free c/a ratio of 1·01; values ranging from 1·01 to 1·0 have been observed by x-rays (Kay 1947), the proportion of material with c/a less than 1·01 increasing when an a.c. field is applied. The possible effects of voids in the ceramic has also been disregarded.

In this type of work spurious results due to experimental conditions and electrical equipment are difficult to avoid when small strains are being observed and integrators of charge are required. Phase shift in amplifiers and bridges must be carefully avoided. Measurements of strain, too, present considerable difficulties because of parasitic modes so frequently found in thin ceramic discs, e.g. 'dishing'; also difficulties are met in measuring specimen thickness in the case of thin discs because

of possible penetration of ' fired-on ' electrodes should the ceramic have less than optimum density.

It is apparent that further research could usefully be devoted to the determination of the relative importance of the various contributing factors in the models proposed, i.e. proportion of $\pm$ domains, proportion of locked domains, proportion of 90° and 180° transitions and contribution of pure piezoelectric effect. Obviously the (D, E) hysteresis loop must play an important part in such analyses; in fact, since the majority of ceramic specimens show strain x proportional to P^2 to a high degree of accuracy, it may only be necessary to understand fully the (D, E) characteristics to understand the striction properties.

Models are again required for preliminary interpretation of the (D, E) characteristics. As an example, consider the simplest case of a single domain having $\pm$ directions only, when the complete set of curves will be as shown in figures 1(*d*), 2(*a*) and 2(*b*), where the straight portions 1–3 and 5–6 in each diagram represent the true piezoelectric component of the strain. When other mechanisms are incorporated into the model, then figure 2(*b*) is modified to figure 3(*a*) for example, if domains rotate altogether but only through 90° steps. Figure 3(*b*) is obtained when all effects are taken into account, and smoothed out because portions of the specimen alter orientation at different parts of the cycle.

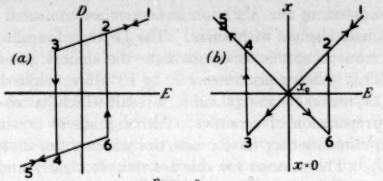


Figure 2

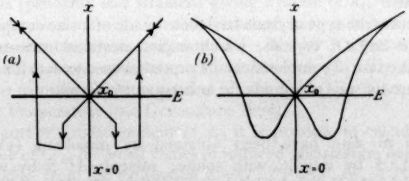


Figure 3.

§ 4. Growth of Single Crystals

Interest in the dielectric and strictive properties of the perovskite ferroelectrics has stimulated attempts to grow crystals of a size suitable for study of atomic confirmations (0·1 to 1·0 mm) and larger crystals for dielectric and electrochemical properties (1·0 mm upwards).

$BaTiO_3$

The earliest recorded success in growing large crystals of size sufficient for all types of electrical measurements was that of Kremers and Williams (1950) and of

von Hippel (1952) who produced $BaTiO_3$ crystals having dimensions rated in inches.

The material was heated to its melting point (1600°c) in a molybdenum furnace whose windings were protected by H_2, while N_2 surrounded the specimen crucible. The crucible material was platinum, which necessitated extensive crucible fabricating equipment. The atmosphere was changed from N_2 to O_2 at 1300°c and held for 15 hours to allow the full stoichiometric oxygen requirements required for the light-tan to water-clear crystals and also to anneal the crystals. No large clear crystals were produced unless 1–2% $SrCO_3$ was added to the original very high purity $BaTiO_3$.

For many investigations rather smaller crystals have been found fairly satisfactory, and these have been produced, of size rarely exceeding 1 mm, by the slow cooling of a solution of $BaCO_3+TiO_2$ or $BaTiO_3$ dissolved in roughly equal weight of $BaCl_2$, from 1100°c to 700°c over a period of days (e.g. Blattner *et al.* 1947). This method has been adopted by the vast majority of workers, but it is clear that the method is not reliable, only a few crystals of sufficient size being obtained in any one experiment. Recently, however, the process has been improved by Remeika (1954) who has produced crystals up to 1·0 cm × 1·0 cm × 0·02 cm in numbers sufficient for many types of investigations.

These large clear ' platey ' crystals are produced by dissolving 30% $BaTiO_3$+ 0·2% Fe_2O_3 in 70% KF. The solution is held at 1150°–1200°c for 8 hours and slowly cooled to 850°–900°c when the still liquid flux is poured off, and the crystals remaining at the bottom of the platinum container are annealed slowly to room temperature and finally leached with water. The Fe_2O_3 is found to reduce the loss by providing electron acceptors to compensate the donors usually imposed by oxygen defects. This addition also lowers T_c by 10°–15°c, while the addition of as much as 2·5% Fe_2O_3 makes the crystal cubic, a result which has also been observed by others in the preparation of ceramics. Although these crystals are ideal for many electrical experiments they have properties which differ in detail from those produced by $BaCl_2$. The reasons for this are not yet clearly understood, and at present comparisons of characteristics are still being made as the newer crystals become available.

No other perovskite type crystals have been made of a size comparable with the largest available $BaTiO_3$ crystals, which makes electrical measurements rather difficult. For the sake of completeness the processes used to date for other materials are recorded here by way of a guide for subsequent researches.

$CaTiO_3$

Best results to date have been obtained by dissolving ($CaO+TiO_2$) in $CaCl_2$: BaCl=1: 5 by weight, with solute: solvent=1: 5 by weight. Slight excess TiO_2 favours clear crystals and sizes up to 1 mm have been obtained containing Ba ions to less than 1 in 10^5 (Kay and Bailey 1955).

$NaNbO_3$

Best results were obtained by Wood (1951) by heating together Nb_2O_5 18·6 g, Na_2CO_3 7·4 g, NaF 12·5 g, for 1 hour at 1500°c in a platinum crucible and slowly cooling. Resulting crystals have size 1 mm × 1 mm × 1 mm.

$LiNbO_3$

Stoichiometrically equal proportions of Li_2CO_3 and Nb_2O_5 are finely ground and mixed. The powder is fired in platinum at 1150°c for 24 hours when crystals up to 1 mm form in the solid.

$KNbO_3$

Matthias *et al.* (1951) claim to have made crystals of up to 1 cm edge. Few details are given, and other workers have failed to attain this size. The only information is that Nb_2O_5 and K_2CO_3 are dissolved in KF and KCl.

Wood (1951) proposed Nb_2O_5 (2·38 by weight)+K_2CO_3(1 by weight) dissolved in KF or KCl at 1000°c and slowly cooled.

$NaTaO_3$, $KTaO_3$

Without details of production methods, both $NaTaO_3$ and $KTaO_3$ are referred to by Matthias (1949 a) and Hulm, Matthias and Long (1950). Crystals up to 1 mm×2 mm×2 mm are claimed, although doubt exists as to the composition of these crystals because of observed variations in Curie temperature.

Extensive experiments in this laboratory have given the best results with $Na_2CO_3 : Na_2B_4O_4 : Ta_2O_5 = 7 : 1 : 4$ heated at 1250°c for 12 hours and cooled over a period of 6 hours. The dimensions obtained have been of the same order as those obtained by Matthias.

§ 5. Striction in Single Crystals

Of the few striction measurements made on crystals (Mason 1948, Jaffe 1950, Rzhanov 1949) those of Caspari and Merz (1951) are probably of greatest interest. They measure the value of d_{23} as a function of temperature, the crystal being maintained as a single domain by a strong polarizing field and the lateral extension measured under static conditions by means of a capacitor-type extensometer. The value obtained for d_{23} is -110×10^{-8} e.s.u., but its variation with temperature (figure 4) does not compare in detail with that calculated from the observed values of spontaneous polarization and dielectric susceptibility $d_{23}=2Q_{23}P_{0}\chi_3$. Dynamic values of d_{23} calculated from resonant and anti-resonant frequency were not absolute but did reveal differences between different crystals, so-called ' good ' crystals (probably less strained) giving a value of d_{23} which more closely follows the theoretical curve. Of some interest here is the large value of d_{23} observed above the transition temperature (-50×10^{-8} c.g.s. at 160°c) and its apparently small decrease with temperature, but one can only assume that this results from the fact that the large applied biasing field (17—20 kv cm^{-1}) raises the effective Curie temperature (see Devonshire 1954).

From the author's measurement of d_{23} it is possible to calculate roughly the equivalent value of Q'_{23} for small values of induced polarization over and above the spontaneous value, from the expression $x_2=Q'_{23}P_3^2=Q'_{23}(P_{03}+P_3)^2$. Now

$$d_{23}=\frac{\partial x_2}{\partial E_3}=2Q'_{23}\frac{\partial P_3}{\partial E_3}(P_{03}+P_3)$$

and assuming P_3 is small compared with P_0, and $\partial P_3/\partial E_3=\epsilon_3/4\pi$ for high permittivity materials in which ϵ does not change to any extent with the strength of field applied, we get $d_{23}=Q'_{23}P_0\epsilon_3/2\pi$. Inserting the value of d_{23} obtained experimentally by Caspari and Merz, and the value of $\epsilon_3=200$ (Merz 1949, Drougard and Young 1954 a) and $P_0=25\times10^{-6}$ coulomb cm^{-2} we find $Q'_{23}=-0{\cdot}46\times10^{-12}$ e.s.u. which compares favourably with the value of $-0{\cdot}49\times10^{-12}$ calculated earlier. The experiments need to be repeated with polarization measurements rather than applied field, and also a measure is needed of effective ϵ_3 over the operating region,

in order to test more conclusively the validity of the simple square law and its temperature dependence. The expected values of d for a polarized single crystal assuming $\epsilon_3=200$ and the Q values obtained earlier are in fact $d_{33}=306\times10^{-8}$, $d_{23}=-109\times10^{-8}$ e.s.u.

Bond, Mason and McSkimin (1951) tested a large single crystal of $BaTiO_3$ by ultrasonic pulse methods and obtained $d_{23}=-310\times10^{-8}$ e.s.u. and $d_{33}=950\times10^{-8}$ e.s.u. (cf. quartz $d_{31}=7\times10^{-8}$ e.s.u.). With the same assumptions concerning the relationship between d and Q' and using their value of $\epsilon_3=1500$ we find $Q'_{23}=-0{\cdot}17\times10^{-12}$ e.s.u. which is one-third the value expected. Here unfortunately the problem is the choice of value of ϵ_3, for the quoted values all seem rather high compared with those of other workers, and it seems apparent that domain effects are complicating these observations.

Hysteresis loops have been discussed by several authors mostly as a passing interest without any detailed explanation of mechanism. Merz (1949) makes the first reference to 'waisted' hysteresis loops obtained in single crystals of $BaTiO_3$. In later papers (Caspari and Merz 1951, Merz 1953, Cross 1953) it is shown that such loops occur only above T_c although the former work indicated otherwise. They can also occur near a ferroelectric–antiferroelectric transition (Shirane *et al.* 1951, Shirane 1952). The explanation lies in the fact that above T_c the material will behave like an ordinary dielectric with a straight line (D, E) curve; higher fields, however, will trigger the spontaneous co-operative mechanism and domains will result giving their own spontaneous strain value and consequent loop as shown idealized in figure 5(*b*), where the rise 1–2 represents the equilibrium value of discontinuous polarization induced at critical field E_c. The value will be somewhat less than that observed at T_c which is about half the room-temperature value of spontaneous polarization $(25\times10^{-6}$ coulomb cm$^{-2})$.

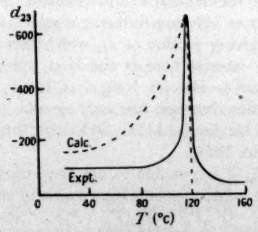


Figure 4

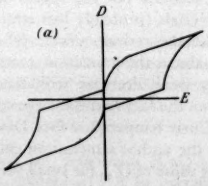


Figure 5 *a*

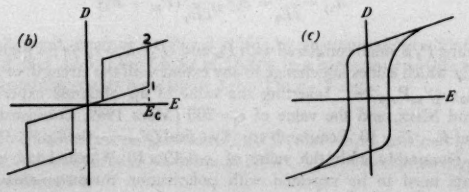


Figure 5

Merz (1952) has contributed to the knowledge of hysteresis loops in single crystals by first establishing directly the presence of 180° domain walls, a feature which previously had not been sufficiently appreciated due to the difficulty of observing them. Correlating these direct optical observations with hysteresis loops (figure 5(*c*)) the author points out that the sharp change of D at low fields results from the fact that it is easier to form completely new lamella-type 180° reversed domains than to extend these lamellae by wall movement (in contrast to the magnetic case), and the more gradual D changes in the opposite quadrants result from boundary wall migrations which require higher fields to excite them and do not commence until some 50% of the crystal has been re-oriented by the formation of fresh nuclei.

Later work by Merz throws further light on the mechanism of domain rotation in crystals. Alternate positive and negative pulses of approximately 10 μsec duration were applied to very thin crystals ($\sim 5 \times 10^{-3}$ cm) at small electrodes, and the duration and amplitude of the current flow through the crystal were recorded as functions of temperature and pulse height. Only 180° domains were operative in these experiments and the conclusions showed that in thin flake crystals 180° rotation occurs by the nucleation of antiparallel domains, probably of a shape long and thin in the polar direction. These extend across the crystal as the field is increased, and exhibit little sideways motion of the walls such as is found in the ferromagnetic case, due to the large wall energy (~ 8 erg cm^{-2}). The reversal process then becomes one of nucleation and propagation of strain points rather than wall motion. Rough calculations also indicate that the 180° wall thickness is only a few lattice units, as is also found in rochelle salt and KH_2PO_4.

Although more important in relation to memory devices (the small amount of wall movement making cross-talk between memory elements less likely) these results will ultimately be of great importance in the complete understanding of hysteresis loops and striction phenomena occurring in ceramics. In this latter connection, observations by Drougard and Young (1954 b) on single crystals indicate that the incremental permittivity decreases at the point where the specimen passes through the region of zero polarization. This suggests that stable arrangements of 'head to tail' domain arrays are readily set up and that this clamping effect reduces the dielectric constant at the point in the hysteresis cycle where in fact the value of $\partial P/\partial E$ is a maximum. This must be borne in mind also when discussing the ageing of ceramics and their very slow relaxation (over many days) into a more stable domain configuration.

Much has yet to be learned to account for the vast differences in coercivity which seem to occur from crystal to crystal selected from nominally the same batch, and between crystals of slightly different chemical composition. An outstanding question of value to the application of piezoelectric materials is whether these differences are essentially chemical (bond resistance to rotation) or physical (impurity-strain, dislocations, free charge layers, etc.).

No single crystals of perovskite materials other than $BaTiO_3$ have been investigated to date.

§ 6. Striction in Ceramic Materials

Of necessity sintered powders are the chief concern of any group interested in practical applications, and not unnaturally work has centred round those having composition closely related to $BaO.TiO_2$. Intensive measurements have been made

on many hundreds of specimens to check the effect of various additions on dielectric properties. Usually the properties measured do not include that of electrostriction but valuable work has been done by measuring ϵ, tan δ etc., of various oxide systems in order to obtain useful dielectric materials.

Apart from these private commercial investigations, very comprehensive surveys have been published by Bunting *et al.* (1947, 1949, 1951) and Shelton *et al.* (1948) on oxide systems including BaO, TiO_2, BeO, MgO, SrO, etc.

The measurements relevant to electrostriction usually made in such survey work include:

(*a*) Variation of dielectric constant with temperature and biasing field at 50, 10^3 and 10^6 c/s.

(*b*) Variation of loss with temperature and biasing field.

(*c*) Hysteresis loops at different temperatures and field strength. Usually these loops are shown on an oscillograph by means of circuits based on that of Sawyer and Tower (1930).

(*d*) Long-term electrical stability.

(*e*) Measurement of d (x/E piezomodulus) and k (coupling factor) by resonance and low-frequency methods.

Commercial firms are naturally unwilling to disclose the actual mixtures that they have found to be useful as a result often of intensive survey work. Many data have been collected, and these have been reduced, for each specimen, to a few reference figures related to loss, resistivity, temperature coefficients, type of (D, E) loop etc., which are to be found for each chemical composition. Associated with these basic measurements much development work has proceeded in parallel: for example, experiments to determine the optimum fabrication and firing conditions to avoid spurious strictive effects due to ' fired in ' strains and porosity, optimum silvering conditions and pre-polarizing techniques.

Few measurements of piezo constant or striction constant have been measured to date, and of these most are indirect. Jaffe (1950) presents the first measurements of the ceramic material in the pre-polarized state, this being an attempt to give the ceramic the properties of a single crystal. His symbolism is slightly different from that used here and has been transposed for simplicity. The volume piezomodulus $d_h=2d_{23}+d_{33}$ was measured by the direct piezoelectric relation between applied stress and resulting charge by placing pre-polarized ceramics inside a hydrostatic pressure cell. Stress was measured by a reference body of a Y-cut crystal of lithium sulphate which has the highest known hydrostatic piezoelectric coefficient $d_h=40{\cdot}5\times10^{-8}$ e.s.u. The results varied considerably with the specimens and covered the range of 6 to 210×10^{-8} e.s.u.

As is so often the case, the results suffer from the lack of direct measurement of some of the other relevant constants. In this case d_{33} was not measured directly and is inferred from the measured d_{23} and d_h. Typical values obtained are $d_{33}=540\times10^{-8}$; $d_{23}=-225\times10^{-8}$; $d_h=93\times10^{-8}$ e.s.u.; $\epsilon_3=1250$ and thickness mode coupling factor $k=0{\cdot}5$, with a remanent polarization (after treatment at 20 kv cm^{-1})of $0{\cdot}24\times10^{-6}$ coulomb cm^{-2}.

It is difficult to convert these measurements to an effective Q value because of the somewhat unreliable estimates of P_0. Inserting the value given produces $Q_{33}=3{\cdot}8\times10^{-11}$ e.s.u. which is twenty times larger than our estimate. However,

using the given measurement of t in the expression $x=tE^2$ for a (Ba.Sr) TiO_3 specimen observed just above its Curie temperature, we obtain from $Q=8\pi^2 t/\epsilon^2$ and the quoted values of $\epsilon_3=3500$, $t_{33}=288\times10^{-9}$ e.s.u. and $t_{23}=108\times10^{-9}$ e.s.u. that $Q_{33}=1{\cdot}86\times10^{-12}$ and $Q_{23}=-0{\cdot}69\times10^{-12}$ e.s.u. which compares reasonably with the single crystal values discussed previously.

Later measurements of Shirane and Sato (1951) provide further results reasonably consistent with the views put forward. These experiments were made with the object of studying the effect of pressure on the dielectric properties of the ceramic. However, use can be made of their conclusions by the relation $Q_{33}=\frac{1}{2}\partial\chi_3/\partial X_3$ etc. where χ is the susceptibility. The equivalence of such an expression can be seen as follows: since $x_h=Q_{hmn}P_mP_n$, $2Q_{hmn}=\partial^2 x_h/\partial P_m\partial P_n$ and from the free energy function $dG_1=-SdT+EdP+xdX$ we have the Maxwell relation for constant T, that $\partial x_h/\partial P_m=\partial E_m/\partial X_h$ where X_h is the stress. Hence

$$2Q_{hmn}=\frac{\partial^2 E}{\partial X_h\partial P_n}=\frac{\partial\chi_{mn}}{\partial X_h}.$$

The values obtained from experiment on (60 Ba. 40 Sr)TiO_3 ceramic are $\partial\chi/\partial X=-0{\cdot}3$ and $+1{\cdot}3\times10^7$ cm^2 kg^{-1} from which can be calculated $Q'_{33}=0{\cdot}82\times10^{-12}$ and $Q'_{23}=-0{\cdot}19\times10^{-12}$ e.s.u. This is slightly lower than the single domain value but is remarkably close considering the dielectric constant measurements in these experiments were made with only 2 volts applied field which would tend to give small Q values. Any domain changes at such low voltages would be of the easier 180° type which contributes polarization change but no strain. The inclusion of 40% Sr for Ba is not expected to alter Q seriously, although it would be of particular interest to see how Q varies with temperature in $BaTiO_3$.

The fact that the measured values of Q for the ceramic discussed so far lie so close to the estimated single crystal values is rather surprising and indicates that in certain types of specimen we can regard the polarized ceramic as a single crystal of spontaneous polarization P whose striction obeys the relation $x=QP^2$ so that $Q=x_0/P_0^2$, and this may be particularly close to the correct situation when the crystallites of the ceramic are small, heavily strained and find domain nucleation difficult.

The only really direct measurements of Q have been made by Gibbs *et al.* (1954). Here the (P_3, x_3) curves have been plotted directly as well as the conventional (D, E) loops at 50 c/s. The results to date indicate:

(i) Pre-polarized ceramics show asymmetric (x, P) curves, as indicated in figure 1(*b*), whose Q value changes little as the pre-polarization is destroyed by application of bigger 50 c/s field. This conforms with the general idea discussed in § 3.

(ii) The value of Q_{33} for specimens of $BaTiO_3$ containing small amounts of impurity varied between 0·30 to $0{\cdot}65\times10^{-12}$ e.s.u. even though the effect of such additives was most marked in the measurement of the peak to peak value of dielectric constant.

(iii) Purer $BaTiO_3$ tended to give (x, P) curves which were not parabolic but rather flattened at the low field portion, becoming steeper with increasing field (cf. figure 1(*c*)). The parabola of best fit corresponds to $Q_{33}=0{\cdot}32\times10^{-12}$ e.s.u. and the portion of the curve at low strain values gives a Q_{33} value several times smaller.

Although the explanation of these low values is by no means certain yet, it probably arises from relative ease of 180° re-orientation which gives little contribution of strain compared with 90° changes which give a much larger strain contribution.

The effective values of Q_{33} observed in the ceramic can vary from a very low value due to piezoelectric contributions (3×10^{-14} e.s.u.) to the maximum value due to domain rotation ($1{\cdot}4\times10^{-12}$ e.s.u.). This can be seen from a rough model in which α domains are considered as contributing to the polarization along the direction of the field, and $1-\alpha$ domains to the polarization against that of the field. Then

$$\text{measured strain} = x_m = \alpha d_{33}E - (1-\alpha)d_{33}E + x'$$

where the first two terms are the piezo contribution, and x' is the actual no field strain value of the cell resolved along the field direction, which value is probably of the form $x' = 0{\cdot}0066P_m^2/P_0^2$, where P_m is the measured polarization along the field direction. This latter expression states that the more random the distribution, the closer to unity will be the c/a ratio, a fact suggested by x-ray measurements of the ceramic.

The measured polarization will be

$$\begin{aligned} P_m &= \alpha(P_0 + P_p) - (1-\alpha)(P_0 - P_p) \\ &= P_0(2\alpha - 1) + P_p \end{aligned}$$

where P_p is the additional polarization due to the field and P_0 is the saturation polarization.

Hence
$$x_m = d_{33}E\left(\frac{P_m - P_p}{P_0}\right) + \frac{0{\cdot}0066P_m^2}{P_0^2}.$$

Now $P_p = P_m\epsilon_c/\epsilon_m$ where ϵ_c and ϵ_m are respectively the dielectric constants along the cell c axis and measured in the experiment.

Therefore
$$x_m = \frac{d_{33}EP_m}{P_0}\left(1 - \frac{\epsilon_c}{\epsilon_m}\right) + \frac{0{\cdot}0066P_m^2}{P_0^2}.$$

Putting $\epsilon = 200$ and $\epsilon_m = 2\times10^4$, which are typical experimental values, we find

$$x_m = 0{\cdot}98\left(\frac{d_{33}4\pi}{P_0\epsilon_m}\right)P_m^2 + \frac{0{\cdot}0066P_m^2}{P_0^2}$$

and putting $d_{33} = 300\times10^{-8}$ e.s.u. and $P_0 = 7{\cdot}5\times10^4$ e.s.u. we find $x_m/P_m^2 = 3\times10^{-14}$ due to the piezoelectric term plus $1{\cdot}17\times10^{-12}$ due to domain strain. The observed value of Q_{33} will then depend on the proportion of each contribution; in the case of very pure specimens 180° rotation predominates at low fields and Q is low, increasing at higher fields as 90° angle domain rotation begins to be more important.

Even at fields as high as 23 kv cm^{-1} (the maximum used) the polarization is only about 50% of that expected from a ceramic ($0{\cdot}83\ P_0$) assuming that the polarization of each crystallite takes up the cube direction aligned most nearly parallel to the field, and the resultant strain is only 25% of that expected for such a polarization. This suggests that about half the polarization is due to 180° rotation which seems a reasonable figure.

Measurements on other materials are infrequently recorded. Gibbs *et al.* (1954) have measured directly different ceramic specimens of (K.Na) $(Nb.Ta)O_3$ and these have been found to give values of $Q_{33}=1{\cdot}76\times10^{-12}$ to 12×10^{-12} e.s.u., which is from 2 to 15 times greater than the highest value recorded for $BaTiO_3$, and it is possible that these materials may have application in the future.

Roberts (1951) measured the piezocoefficient of ceramic discs of $PbZrO_3$ after subjecting them to 2·5 kv cm^{-1} while cooling through the crystal transition at 220°c. The result is small: $d_{33}=3\times10^{-9}$ e.s.u., i.e. just 0·05 the value of d_{31} for quartz. The reason for this is obviously the antiferroelectric nature of this material (Shirane *et al.* 1951, Shirane 1952). Any piezo effect resulting from superimposing an electric field on the spontaneous antiparallel dipole alignment must be zero, but in fact the antiparallel array occurs in the *ab* plane of the cell while the symmetry allows a polar *c* direction (Sawaguchi *et al.* 1951) and this must be the contributing direction. No evidence is available concerning details of the structure but it is reasonable to assume that the polarity of the *c* direction was not reversed by the field during these experiments, otherwise a symmetric parabolic type (x, E) curve would have been observed.

So far we have ignored materials which have other than the perovskite structure, but if ferroelectricity occurs in perovskites as a result of accidentally critical spacings between highly polarizable ions, it is only reasonable to expect that very many other solids will exhibit similar phenomena. Several of the ilmenite type materials have been recorded as showing interesting dielectric properties, e.g. $LiNbO_3$ and $LiTaO_3$ (Matthias 1949 b), but so far none have shown very useful piezoelectric characteristics. On the other hand, complete departure from the ABO_3 perovskite or ilmenite structures has revealed an interesting material in $Pb(NbO_3)_2$, a material which was selected for synthesis in view of the high value of the polarizability/atomic-radius ratios for its constituent atoms, indicated also by a refractive index of 2·5. This material has been prepared in powder form and investigated by Goodman (1952); its interesting properties may be best described by comparing results obtained from a ceramic of this material with ceramic $BaTiO_3$.

	$Pb(NbO_3)_2$	$BaTiO_3$
Density (g cm^{-3})	6·33	5·7
Refractive index	2·5	2·5
Dielectric constant	280	1700
d_{33} (c.g.s.)	248×10^{-8}	590×10^{-8}
Young's modulus (dyn cm^{-2})	$6{\cdot}2\times10^{11}$	$9{\cdot}4\times10^{11}$
Coupling coefficient (thickness mode)	0·26	0·21
Useful temperature range (°c)	−200 to +500	0−100

Its structure is by no means as simple as the ABO_3 compounds, and no details are yet available regarding the mechanisms of polarization or domain orientation.

Hysteresis loops of ceramic materials have excited considerable interest. Kambe *et al.* (1953), realizing that little was known regarding the relative contributions to the polarization of domain movement and other effects, investigated the hysteresis loops of $BaTiO_3$ ceramic by means of $1{\cdot}5\times10^{-4}$ sec pulsed a.c. fields at frequencies from 100 kc/s to 7·5 Mc/s. The pulsing was solely to eliminate spurious effects due to dielectric heating.

Pulsed techniques gave more rounded ' points ' and lower values of ϵ (peak to peak average). For a given field the maximum P and maximum width decreased with increasing frequency. By altering the pulse length at 10 kc/s it was then shown that the first few cycles produced a slightly ' waisted ' loop which gradually fills out and develops saturation points in a period of 15–20 seconds. From their results the authors conclude that two relaxation processes have been observed, one extremely rapid process being completed within 3×10^{-5} second and another associated with the widening and development of saturation points in a time up to several seconds. This latter was tentatively associated with domain 'redistribution' occurring before cyclic conditions are firmly established.

While these results give pointers to future effort, they only contribute generalizations which are not entirely contrary to expectations, and which are demonstrated more simply by the observations of Gibbs *et al.* (1954). In these experiments (at 50 c/s) a field equal to about half the saturation field is suddenly increased. For a short period thereafter the hysteresis loop shows a memory of its previous shape (as shown dotted in figure 6) which gradually fades out within a period of about a

Figure 6

second. At the same time the point becomes sharper and the loop length increases slightly in both length and width. The effect varies in magnitude between different specimens, and presumably is associated with the longer period change observed by the Japanese workers.

A great deal of additional work is necessary to interpret the results but one may tentatively assume that this delay in achieving the cyclic condition is attributable to some blocking mechanism of domain wall movement, i.e. resistance to the motion of the strained leading points of lamellar domains, or resistance to lamellar nucleation. The blocking points may be strains such as dislocations, impurity atoms, crystallite boundaries, or charge concentrations. If dislocations were the cause, for example, then the domain motion could be considered as sweeping up these dislocations into piles at the extremities of the boundary movement. When the field is increased the boundaries cannot readily cross over them so that for some time the loop tends to retain some of its original characteristics. The kink occurs where domain walls are crossing dislocation piles which need some time to shift, depending on their natural temperature-dependent mobility and the excess field applied. Such a mechanism would also account for the higher values of ϵ obtained after the application of a strong a.c. field to freshly fired ceramic (Gibbs *et al.* 1954). Measurements at different temperatures will no doubt provide further information.

A third, and even longer period, effect is recorded by Mason (1954) who gives details of the ageing of $BaTiO_3$ ceramic transducers which show variations over

periods of hundreds of days of 15–20% in dielectric constant, 5–25% in piezo constant and 100–300% in electrical quality factor. In this case careful measurements at different temperatures indicated a domain relaxation process having an activation energy of 19 kcal mole^{-1}. This was associated with the relaxation of 45° domain boundaries from the asymmetrical conditions of dipoles in a crystallite imposed by the cooling stresses, to a more regular equivolume array. The relaxation is effected by small thermally activated steps of units of seven unit cells. Such investigations are particularly important for the stabilization of commercially produced transducers, and it is shown that the addition of Ca and Pb in place of Ba reduces the ageing required before the subsequent drift of properties is reduced to acceptable proportions. The exact mechanism of this stabilization is not yet clear.

It is not intended to give, in this account, details of present trends of practical application. A useful introductory survey of the application to oxide-type transducers is given by Bradfield (1954) and Mason and Wick (1954).

We have seen how the perovskite-type materials are both electrostrictive and piezostrictive in behaviour depending on temperature, and how it is sometimes more convenient to maintain an electrostrictive description even when the individual components can only be described as piezoelectric units. Although the constants for the single domain, at first sight, bear no relation to the state of affairs expected in the ceramic, measurements show very close similarities except for a general tendency for lower constants in the ceramic indicating that 180° domain reversals play a significant role in the process.

In conclusion, it seems that the general transducer behaviour of $BaTiO_3$ materials is well understood in broad outline but many details have yet to be unravelled. To this end, further researches are still required into improved methods of growth of crystals of size suitable for visual domain observations and electrical measurements, not only of $BaTiO_3$ alone, but also $BaTiO_3$ plus controlled additives and of other materials which show effects of reasonable magnitude. Also much more is yet to be understood of the mechanisms of domain propagation, resistance and relaxation, so that more detailed pictures of the behaviour of single crystals and the technically important ceramics can be obtained.

References

Anderson, J. R., 1952, *Elect. Engng*, 916.

Blattner, H., Matthias, B. T., and Merz, W., 1947, *Helv. Phys. Acta.*, **12**, 35.

Bond, W. L., Mason, W. P., and McSkimin, H. J., 1951, *Phys. Rev.*, **82**, 442.

Bradfield, G., 1954, *Acustica*, **4**, 200.

Brown, C. S., Kell, R. C., Thomas, L. A., Wooster, N., and Wooster, W. A., 1952, *Miner. Mag.*, **29**, 858.

Bunting, E. N., Shelton, C. R., and Creamer, A. S., 1947, *Bur. Stand. J. Res., Wash.*, **38**, 337; 1949, *Ibid.*, **43**, 237.

Bunting, E. N., Shelton, C. R., Creamer, A. S., and Jaffe, B., 1951, *Bur. Stand. J. Res., Wash.*, **47**, 15.

Cady, W. G., 1946, *Piezoelectricity* (New York: McGraw Hill).

Caspari, M. E., and Merz, W. J., 1951, *Phys. Rev.*, **80**, 1082.

Cross, L. E., 1953, *Phil. Mag.*, **44**, 1161.

Devonshire, A. F., 1949, *Phil. Mag.*, **40**, 1040; 1951, *Ibid.*, **42**, 1065; 1954, *Advances in Physics* (*Phil. Mag.* Suppl.), **3**, 85.

Drougard, M. E., and Young, D. R., 1954 a, *Phys. Rev.*, **94**, 1561; 1954 b, **95**, 1152.

EUCKEN, A., and BÜCHNER, A., 1934, *Z. Phys. Chem.*, **B, 27,** 321.
GIBBS, D. F. *et al.*, 1954, *Research Reports* to Signals Research and Development Establishment, Ministry of Supply.
GOODMAN, G., 1952, *Report No. RL*/79.3. General Electric Research Laboratory, N.Y.
VON HIPPEL, A., 1950, *Rev. Mod. Phys.*, **22,** 221; 1952, *Tech. Rep.* 51, (M.I.T.).
HULM, J. K., MATTHIAS, B. T., and LONG, E. A., 1950, *Phys. Rev.*, **79,** 885.
JACKSON, W., 1950, *Proc. Instn Elect. Engrs*, **97,** 285.
JAFFE, H., 1950, *Indian Engng Chem.*, **42,** 264.
KAMBE, K., *et al.*, 1953, *J. Phys. Soc. Japan*, 8, 9 and 15.
KAY, H. F., 1947, *Acta. Cryst., Camb.*, **1,** 229 (*Thesis*, University of Manchester).
KAY, H. F., and VOUSDEN, P., 1949, *Phil. Mag.*, **40,** 1019.
KAY, H. F., and BAILEY, P., 1955, in preparation. Also see BAILEY, *Thesis*, University of Bristol, 1952.
KREMERS and WILLIAMS, 1950, U.S. Contract No. N-36-039-Sc-38180 (Harshaw Chemical Co. Report).
LUTTINGER, J. M., and TISZA, L., 1946, *Phys. Rev.*, **70,** 954.
MASON, W. P., 1948, *Phys. Rev.*, **74,** 1134; 1951 a, *Piezoelectric Crystals and their Applications to Ultrasonics* (New York: van Nostrand); 1951 b, *Phys. Rev.*, **82,** 442; 1954, *Acustica*, **4,** 200.
MASON, W. P., and WICK, R. F., 1954, *Proc. Inst. Radio Engrs, N.Y.*, **42,** 1606.
MATTHIAS, B. T., 1949 a, *Phys. Rev.*, **75,** 1771; 1949 b, *Ibid.*, **76,** 1886.
MATTHIAS, B. T., and REMEIKA, J. P., 1951, *Phys. Rev.*, **82,** 727.
MEGAW, H. D., 1952, *Acta Cryst., Camb.*, **5,** 739; 1954, *Ibid.*, **7,** 187.
MERZ, W. J., 1949, *Phys. Rev.*, **76,** 1221; 1952, *Ibid.*, **88,** 421; 1953, *Ibid.*, **91,** 513.
REMEIKA, J. P., 1954, *J. Amer. Chem. Soc.*, **76,** 940.
ROBERTS, S., 1951, *Phys. Rev.*, **83,** 1078.
RZHANOV, A. V., 1949, *J. Exp. Theor. Phys., U.S.S.R.*, **19,** 502.
SAUER, J. A., 1940, *Phys. Rev.*, **27,** 142.
SAWAGUCHI, E., MANIWA, H., and HOSHINO, S., 1951, *Phys. Rev.*, **83,** 1078.
SAWYER, C. B., and TOWER, C. H., 1930, *Phys. Rev.*, **35,** 269.
SHELTON, C. R., CREAMER, A. S., and BUNTING, E. N., 1948, *Bur. Stand. J. Res., Wash.*, **41,** 17.
SHIRANE, G., 1952, *Phys. Rev.*, **86,** 219.
SHIRANE, G., HOSHINO, S., and SUZUKI, K., 1950, *J. Phys. Soc. Japan*, **5,** 453.
SHIRANE, G., and SATO, K., 1951, *J. Phys. Soc. Japan*, **6,** 20.
SHIRANE, G., SAWAGUCHI, E., and TAKAGI, Y., 1951, *Phys. Rev.*, **84,** 476.
SLATER, J. C., 1950, *Phys. Rev.*, **78,** 748.
THOMAS, L. A., WOOSTER, N., and WOOSTER, W. A., 1949, *Disc. Faraday Soc.*, No. 5, 341.
VIGOUREUX, P., 1939, *Quartz Oscillators and their Applications* (London: H. M. Stationery Office); 1950, *Ultrasonics* (London: Chapman and Hall).
WOOD, E. A., 1951, *Acta. Cryst., Camb.*, **4,** 353.

MAGNETIC COOLING

By E. AMBLER and R. P. HUDSON

Cryogenic Physics Laboratory, National Bureau of Standards, Washington 25, D.C.

CONTENTS

	PAGE
§ 1. Introduction	252
§ 2. Cooling by adiabatic demagnetization	254
2.1. Theory	254
(*a*) General	254
(*b*) Effect of crystalline electric field	256
(*c*) Interaction of magnetic ions	258
(*d*) Hyperfine splitting	260
2.2. Experimental	261
§ 3. The determination of temperatures below 1°K	264
3.1. The 'magnetic scale'	264
3.2. Calorimetric methods	265
3.3. Magnetic methods	267
3.4. The 'theoretical method'	268
3.5. Secondary thermometers	268
3.6. The results	269
3.7. Note on discrepancies	274
§ 4. Properties of paramagnetic salts below 1°K	274
4.1. Susceptibility	274
4.2. Relaxation and hysteresis phenomena	274
4.3. Specific heat anomalies	276
4.4. Free radicals and F-centres	277
4.5. Antiferromagnetism	278
4.6. Thermal conductivity	279
§ 5. Thermal insulation and thermal contact below 1°K	280
5.1. Thermal insulation	280
5.2. Thermal contact	282
§ 6. Experiments below 1°K	284
6.1. Properties of metals	284
(*a*) Electrical conductivity	284
(*b*) Superconductivity	285
(*c*) Thermal conductivity	286
6.2. Properties of helium	287
(*a*) Liquid helium	287
(*b*) Solid helium	289
(*c*) Helium 3	290
§ 7. Further developments in magnetic cooling	292
7.1. Two-stage demagnetization	292
7.2. Nuclear demagnetization	292
7.3. Continuous cooling cycles	294
7.4. The use of anisotropic paramagnetic salts	295
§ 8. Nuclear orientation	296
Acknowledgments	299
References	299

Abstract. After a historical introduction the general principles of the magnetic cooling method are outlined. First, the case of the ideal paramagnetic salt is treated, and this is followed by a discussion of the electric and magnetic interaction effects which cause departures from ideal behaviour, and set a lower limit to the temperatures which can be reached upon

adiabatic demagnetization. A description of experimental techniques and procedures includes the subjects of cryostat and magnet design. The methods which have been used to establish the absolute temperature in the range below 1°K are surveyed and a summary of the results obtained is presented. Apart from their usefulness as coolants and thermometers, paramagnetic salts have intrinsically interesting thermal and magnetic properties, e.g. specific heat anomalies and hysteresis effects, associated with the internal ordering which proceeds as the temperature approaches the absolute zero. These properties are discussed in a separate section; another deals with the properties of other substances in the region below 1°K—mainly metals and liquid helium. A section is also included on the difficult problems associated with providing adequate thermal insulation for a low temperature system on the one hand, and thermal contact within such a system on the other. The comparatively new developments of cascade demagnetizations and continuous cooling cycles are described. Finally, the principles involved in a recent striking application of the magnetic cooling method, viz. the spatial orientation of atomic nuclei, are outlined.

§ 1. Introduction

The temperature region below 1°K constitutes a wide and relatively new field for physical research and may only be entered, effectively, by means of the process of ' magnetic cooling '. It was less than thirty years ago, in 1926, that Debye (1926) and Giauque (1927) suggested, independently of each other, that the reversible temperature effects connected with the magnetization of certain paramagnetic salts might be used for producing temperatures well below those obtainable by means of liquid helium. The first experiments with this method were carried out in 1933 (Giauque and MacDougall 1933) and since then magnetic cooling has become an established cryogenic technique, both for studying the properties of the paramagnetic salts used as cooling agents and for carrying out experiments on other substances in this new temperature region.

Langevin (1905) was the first to point out that changes in the magnetization of paramagnetic substances must generally be accompanied by reversible temperature changes. Langevin's theory was based on the assumption of an assembly of non-interacting magnetic dipoles which, in the absence of an external magnetic field, are pointed at random. Upon application of a magnetic field H the energy of a given dipole depends upon its orientation with respect to the field direction and a statistical equilibrium is produced, depending on the temperature T in accordance with the Boltzmann statistics. It followed from this theory that the magnetic moment M should depend only upon the ratio H/T, and for small values of H/T this leads to Curie's law, according to which the paramagnetic mass susceptibility, that is, the magnetic moment per unit mass divided by H, is inversely proportional to the absolute temperature. (The symbol χ will be used to denote the volume susceptibility i.e. magnetic moment per unit volume I divided by H; χ also obeys Curie's law provided that one may neglect thermal expansion—as is the case at low temperatures.)

If pressure and volume effects are neglected one may write down the expression for the enthalpy E as $E=U-HM$, where U is the internal energy. The condition for dE $(=TdS-MdH)$ to be a total differential gives

$$\left(\frac{\partial T}{\partial H}\right)_S = -\left(\frac{\partial M}{\partial S}\right)_H \qquad \ldots\ldots(1)$$

$$= -\frac{T}{C_H}\left(\frac{\partial M}{\partial T}\right)_H$$

where $C_H = T(\partial S/\partial T)_H$ is the specific heat in constant magnetic field. If $(\partial M/\partial T)_H$ is negative, as is the case with a substance obeying Curie's law, isentropic magnetization will be accompanied by heating and demagnetization by cooling.

On the kinetic picture, any degree of alignment of the dipoles with the magnetic field results in a decrease in potential energy. Since the internal energy is, by definition of an ideal paramagnetic substance, independent of the relative orientation of the dipoles, the energy of magnetization appears in the form of heat. Langevin pointed out that the effect is bound to be small at ordinary temperatures, where $\partial M/\partial T$ is small and the heat capacity large.

The position altered completely with the discovery by Woltjer and Kamerlingh Onnes (1923) that some paramagnetic salts, e.g. gadolinium sulphate, behaved according to Langevin's theory even at the temperature of liquid helium. The conditions for appreciable magnetocaloric effects, namely small specific heats and high values of $\partial M/\partial T$, could now be satisfied and it was the recognition of this fact that led Debye and Giauque to put forward their suggestions for the magnetic cooling method.

In their first experiment Giauque and MacDougall, using gadolinium sulphate, reached 0·53°K, starting from 3·4°K and 8 kilo-oersteds, and subsequently 0·25°K with the same initial field but demagnetizing from 1·5°K. The final temperatures were estimated by measuring the magnetic susceptibility and extrapolating the Curie relationship. Giauque's experiments were followed very soon by those of de Haas, Wiersma and Kramers (1933 a) who used cerium fluoride in their first experiments and reached 0·27°K starting from 1·26°K and 30 kilo-oersteds. Here again the temperature was estimated from the susceptibility, which was determined by a balance method. The final field after demagnetization had to be, therefore, of the order of 1000 oersteds and this was one of the reasons why, in spite of the high initial field, lower temperatures were not reached. Better results were obtained by using cerium and dysprosium ethyl sulphates and potassium chromic alum (de Haas, Wiersma and Kramers 1933 b).

Kurti and Simon (1934) carried out their first adiabatic demagnetization experiments with manganous ammonium sulphate, and by cooling other substances in contact with the paramagnetic salt these authors found some new superconductors with transition points below 1°K (1934, 1935 a, b).

The determination of the absolute temperature scale below 1°K soon became one of the chief preoccupations of workers in this field. The first experiments using a purely thermodynamic method were made by Giauque and MacDougall (1935 a) on gadolinium phosphomolybdate and by Kurti and Simon (1935 c) on ferric ammonium alum, while de Haas and Wiersma (1936) estimated the temperatures reached with caesium titanium alum by a 'magnetic method'. Knowledge of the absolute temperature scale made it possible to obtain more detailed information on the nature and magnitude of the interactions in paramagnetic salts and to discuss the experimental results in the light of theoretical work by Van Vleck (1937) and by Hebb and Purcell (1937).

Experiments on thermal conductivities and relaxation phenomena in paramagnetic salts below 1°K were begun in 1936 (Kurti, Rollin and Simon 1936, Cooke and Hull 1937). The period up to 1939 also saw the first experiments on liquid helium below 1°K (Kurti and Simon 1938 a), extension of the work on the properties of paramagnetic salts (Casimir, de Haas and de Klerk 1939 a, b), on heat transfer and relaxation effects (Shire and Allen 1938, Shire and Barkla 1939), and

the investigation of the occurrence of co-operative phenomena in paramagnetic salts at very low temperatures (Kurti, Lainé, Rollin and Simon 1936 a, Kurti, Lainé and Simon 1937).

After 1946, work with magnetic cooling restarted with great intensity and at present more than a dozen laboratories throughout the world are known to be actually employing this technique.

§ 2. Cooling by Adiabatic Demagnetization

2.1. *Theory*

(*a*) *General.*

The aim of the theory is the prediction of the various thermal and magnetic properties of paramagnetic salts, such as entropy S, and magnetization M, and their correlation with the experimental data. All these functions can be readily derived from the partition function Z of the assembly of paramagnetic ions contained in the salt. The well-known expressions of statistical mechanics give

$$S=k\frac{\partial}{\partial T}(T\log Z) \qquad \ldots\ldots(2)$$

and

$$M=kT\frac{\partial}{\partial H_e}(\log Z) \qquad \ldots\ldots(3)$$

where H_e is the ' external ' magnetic field, i.e. the value which would exist in the absence of the specimen, and k Boltzmann's constant. In the limiting case of very elongated specimens placed parallel to the field H_e is equal to H, the magnetic field within the specimen, which appears in the definition of the susceptibility, i.e. $\chi=I/H$.

If the dipoles can be regarded as identical and independent of each other, the partition function can be written $Z=z^N$ where N is the number of dipoles and z is the partition function for a single dipole. The entropy is then given by $S/R=\partial(T\ln z)/\partial T$ where S/R is a dimensionless quantity, not depending on the quantities of material considered, and R is the gas constant.

In this report we shall only be concerned with salts of the iron group and of the rare earth group. Their paramagnetism is due to the magnetic moment associated with the electrons in the incompletely filled shell of the metallic ion. It is advantageous to begin a discussion of the theory with a treatment of the case of free ions, for which the Russell–Saunders coupling scheme applies. At liquid helium temperatures only the lowest of the multiplets need be considered. It is characterized by a quantum number J which determines the resultant angular momentum of the electrons, and a quantum number M_J which determines the component of the angular momentum along a chosen axis. M_J can only take one of the $2J+1$ discrete values $J, J-1, \ldots -J$, the corresponding components of the magnetic moment along the axis being $gM_J\beta$, where β is the Bohr magneton and g the spectroscopic splitting factor. In zero external field these magnetic substates are degenerate, the partition function $z=2J+1$ and $S/R=\ln 2J+1$.

In a magnetic field H one has

$$z=\sum_{M_J=-J}^{M_J=+J}\exp\left[-gM_J\beta H/kT\right]$$

and by suitable differentiation one obtains

$$M=M_\infty\left[\frac{2J+1}{2J}\coth\frac{2J+1a}{2J}-\frac{1}{2J}\coth\frac{a}{2J}\right] \qquad \ldots\ldots(4)$$

$$S=R\ln(2J+1)+\int_0^H\left(\frac{\partial M}{\partial T}\right)_H dH \qquad \ldots\ldots(5)$$

where $M_\infty=NgJ\beta$ is the saturation moment. The expression on the right-hand side of equation (4) is the familiar Brillouin function. These functions have been tabulated by Hull and Hull (1941) for $J=1/2$, $3/2$, $5/2$, by Giauque, Stout, Egan and Clark (1941) for $J=7/2$, and by Schmid and Smart (1954) for $J=1/2, 1, \ldots, 9/2$. Although they correspond to the cases of completely free ions, it turns out that they are of use for calculating the entropy of paramagnetic salts used in the magnetic cooling process under the starting conditions, i.e. at temperatures of about 1°K, if the appropriate quantum number is used for J—see § 2.1 (*b*).

For the majority of salts used in the magnetic cooling process, the lattice entropy is quite negligible compared with the 'magnetic entropy' even at the initial temperature, which nowadays is usually about 1°K. Hence a description of the process which equates the increase of magnetic entropy to the decrease of thermal entropy of the lattice no longer applies. One must rather view absolute temperature in its statistical aspect, i.e. as a parameter in the partition function determining the distribution among the different states. Then upon adiabatic demagnetization the temperature will fall until the various internal constraining forces produce the same degree of order in zero field as that produced by the external magnetic field at the starting temperature. Two different types of constraining forces together bring about the reduction in entropy in zero field; one is the effect of the electric field due to other ions distributed on crystal sites round the paramagnetic ion, and the other the effect of direct interactions between the magnetic ions. Two curves, ABO and ACO, have been drawn in figure 1 to represent the behaviour of two

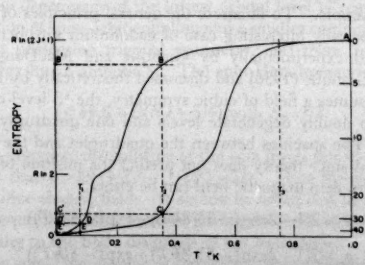


Figure 1. Entropy as a function of temperature for two hypothetical paramagnetic salts.

hypothetical salts, showing an entropy reduction to R ln 2 due to the effect of the crystalline electric field (see § 2.1 (*c*)), followed by the further reduction due to interactions. The magnitude of the interactions is smaller in the salt represented by ABO. The numbers marked along the 1°K ordinate give typical values of the magnetic field in kilo-oersteds (corresponding roughly to the case of $J=5/2$) which must be applied at that temperature in order to obtain the corresponding entropy reductions.

With a given magnetic field it is possible to reach a lower temperature with the salt wherein the interactions are smaller. Furthermore, since this salt has a bigger heat content ($\int TdS$) in the low temperature range T_1 to T_2—compare areas BB'D'D and CC'E'E—it would be possible to *maintain* the temperature for a longer period of time. Similarly, the salt represented by curve ACO would be most suitable for work in the range T_2 to T_3, where its entropy is falling most rapidly and where its specific heat is consequently greatest.

(*b*) *Effect of crystalline electric field.*

In the salts most commonly used each metallic ion is surrounded by six water molecules situated at the corners of an octahedron. When this octahedron is regular the electric field at the central ion has cubic symmetry; it may be distorted, however, to give components of lower symmetry. Bethe (1929) was the first to treat the problem of the splitting of a degenerate level through the effect of this electric field and showed how the splitting depends upon the nature of the term and the symmetry of the electric field. The theory has been applied in detail to various paramagnetic salts by Kramers (1930), Penney and Schlapp (1932), Van Vleck and Penney (1934) and, with particular respect to the magnetic cooling method, by Hebb and Purcell (1937).

The splittings may be quite small ($\sim 0{\cdot}1$ cm^{-1}) in the case of ions in S states, e.g. Fe^{3+}, or they may be very large (~ 1000 cm^{-1}) in cases where there is a resultant orbital angular momentum, e.g. Cr^{3+} and Co^{2+}. In either case, however, the result usually is that we have a low-lying group of levels so well separated in energy from higher levels that they only are populated at the temperatures used at the start of adiabatic demagnetization experiments. This low-lying group of levels cannot usually be characterized by the same quantum numbers which were used to specify states of the free ion, and their behaviour magnetically may be quite complicated.

When the splitting pattern is known, the task of writing down the partition function is fairly simple. To illustrate the general principles of the method, one may take the historically interesting case of gadolinium nitrobenzene sulphonate, which was studied experimentally by Giauque and MacDougall (1935 a) and MacDougall and Giauque (1936) and discussed theoretically by Hebb and Purcell (1937). If one assumes a field of cubic symmetry, the ^8S level of the gadolinium ion splits into two doubly degenerate levels and one quadruply degenerate level lying between.† The spacings between the quadruplet and the two doublets are in the ratio 5:3 but since theory does not predict the position of the spacings the partition function in zero magnetic field can be either,

$$z_c = 2(1+2\exp(-3\delta/8kT)+\exp(-\delta/kT)) \qquad \ldots\ldots(6a)$$

or

$$z_c = 2(1+2\exp(-5\delta/8kT)+\exp(-\delta/kT)) \qquad \ldots\ldots(6b)$$

where δ is the overall energy separation. Hebb and Purcell found that the experimentally determined entropy–temperature curve agrees better with the theoretical curve calculated from (6*a*); δ/k was found to be $1{\cdot}4$ degrees.

In order to calculate the susceptibility it is necessary to know how the levels split when a small magnetic field is applied. Hebb and Purcell summarize this

† Paramagnetic resonance experiments on other gadolinium salts (Bleaney, Elliott Scovil and Trenam 1951, Trenam 1953, Bogle and Heine 1954) indicate, however, that the correct splitting pattern is one of four doublets.

information by giving the magnetic moment matrix in a representation which makes the energy diagonal. The elements of this matrix are substituted into the standard formula for the magnetic susceptibility (see for example Van Vleck 1932). They find $\chi=\gamma\tau/3T$ where $\tau=Ng^2\beta^2J(J+1)/k$ with $g=2$, $J=S=7/2$ and N the number of ions per cm^3; γ is a function of δ/kT which we shall not quote here.

Until about 1946, adiabatic demagnetization experiments were the most important source of information for these small energy splittings occurring in paramagnetic salts. Recently, however, microwave spectroscopy has provided a much more powerful tool for determining splitting patterns and splitting constants, and for predicting entropy reductions caused by the crystalline electric field. It is all the more useful in that it allows one to examine whether, as is often the case, there are different sets of ions in the unit cell of the crystal which have different axes of symmetry and different splitting patterns. The results of microwave spectroscopy are usually expressed by giving the values of certain constants in a ' spin Hamiltonian' (Abragam and Pryce 1951) which is a function of an angular momentum operator S. The value of S is obtained by setting $2S+1$ equal to the total number of levels (including degeneracy) in the lowest-lying group, and may or may not be the same as the actual value of the electron spin of the free ion. The energies of the states of the paramagnetic ion (as a function of an externally applied magnetic field) are the same as the eigenvalues of the spin-Hamiltonian. An excellent detailed discussion, including a derivation of the spin-Hamiltonian in many special cases, has been given recently by Bleaney and Stevens (1953). Here it will suffice to quote one example by way of illustration, namely that of the Cr^{3+} ion ($L=3$, $S=3/2$) in the alums. The crystalline electric field has predominantly cubic symmetry and splits the seven-fold orbital level into two triplets, and a singlet lying about 10^4 cm^{-1} below the nearest triplet. Under these conditions there is very little contribution to the susceptibility from orbital motion (the orbital angular momentum is said to be ' quenched ') and the g-value is only very slightly less that the free spin value of 2. The four-fold spin degeneracy of the lowest orbital level is removed, however, by the combined action of spin–orbit coupling and a smaller component of the crystalline electric field possessing trigonal symmetry, and there result two doublets separated in energy by an amount δ. The spin-Hamiltonian is given as

$$g\beta\,\mathbf{H}.\mathbf{S}+D(S_z^2-5/4) \text{ with } S=3/2.$$

The second term describes the splitting of the spin levels due to the crystalline field and one can see that $\delta=D\{(3/2)^2-(1/2)^2\}=2D$. Here the z-axis is the trigonal axis of the crystalline electric field. It should be noted that in the alums there are four sets of non-equivalent ions, each set having its z-axis in one of the four directions corresponding to the body diagonals of the basic cube.

In general the spin-Hamiltonian is somewhat more complicated; for example, the g-factor may be anisotropic, or the crystal field may have lower symmetry so that terms in S_x^2 and S_y^2 have to be introduced. However, one can always construct the energy matrix from it and, in principle, obtain the partition function and hence derive expressions for S/R and χ.

In adiabatic demagnetization experiments one is generally interested in (*a*) the expression for the entropy of the specimen under the initial magnetizing conditions at temperatures around 1°K, and (*b*) the expressions for the entropy and susceptibility in zero field at temperatures below 1°K. The former usually involves computing a

small correction to be applied to the expression given in equation (5). It has been done for the important example of Cr^{3+} in the alums for the case of a single crystal with the magnetic field applied along the direction of a cube edge, i.e. when the magnetic field makes the same angle with each of the trigonal axes of the four sets of ions (Hudson 1952) and for the case of a powdered specimen (Daniels and Kurti 1954). The case of copper potassium sulphate in which the Cu^{2+} ion has a slightly anisotropic g-factor has been treated by Garrett (1950 a).

In principle, the expressions for the entropy in zero magnetic field and the susceptibility are obtainable directly from the spin-Hamiltonian. In practice, however, only the values in the higher temperature region turn out to be useful on account of the complicating effects of the direct interactions between ions (discussed in the next section). In the high temperature region the susceptibility may be expressed in a power series in $1/T$ by, for example, first expanding the partition function in the same way. The first term in the expansion gives the Curie constant and the second a contribution to the Curie–Weiss constant Δ, as may be seen by writing the Curie–Weiss law,

$$\chi=\frac{c}{T-\Delta}=\frac{c}{T}+\frac{c\Delta}{T^2}+\ldots.$$

It should be pointed out that in substances having an isotropic g-factor the average value of Δ vanishes for powdered specimens, and also for single crystals of substances like the alums where the non-equivalent ions have their axes arranged in cubic symmetry (Penney and Schlapp 1932). For example, the first part of the expansion for the susceptibility of the Cr^{3+} ion in the alums is, considering only the effect of the crystalline electric field,

$$\chi=\frac{5N\beta^2}{k}\left(\frac{1}{T}-\frac{\delta^2}{30T^3}\right).$$

(*c*) *Interaction of magnetic ions.*

It follows from a theorem by Kramers (1930) that for paramagnetic ions with an odd number of electrons (this condition is fulfilled by all those used in magnetic cooling) the crystalline electric field cannot entirely remove the degeneracy. The remaining (two-fold) degeneracy can only be removed by interaction between the ions themselves. The forces concerned here may be either of the magnetic dipole type, the exchange type, or both. One now has to deal with an N-body problem, N being the number of paramagnetic ions in the crystal as a whole, and there will be a band of 2^N energy levels. The crystal in its lowest state could be either ferromagnetic with all the dipoles pointing in the same direction, thus giving rise to a spontaneous magnetization, or antiferromagnetic, consisting of some ordered array of mutually compensating dipoles. The behaviour of paramagnetic salts in this co-operative region is not too well understood; discussion on this aspect will be postponed until § 4.5. For the present it is sufficient to say qualitatively that in most cases the final reduction of the entropy to zero will be brought about by a co-operative action of the dipoles and will be steeper than is the case for the effect of the crystalline field. This means that the temperature remains substantially constant over a considerable range of entropy or, in other words, for a wide range of values of the magnetic field used for the adiabatic demagnetization process.

The effect of these interactions in the higher temperature region is quite well understood, however, and it is this aspect with which this section will be mainly

concerned. There are three main approaches. In the Lorentz method one represents the magnetic interaction of the dipoles by a local field nI, where $n=4\pi/3$ for a cubic lattice. This approach is formally the same as the Weiss molecular field treatment of ferromagnetism, and predicts ferromagnetism at low temperatures except for the case of a spherical specimen. As was pointed out by Onsager (1936) however, the Lorentz theory applies to a system of induced, as distinct from permanent, dipoles. Onsager, using classical magnetostatics and treating the surroundings of a given dipole as a continuum, attempted a more rigorous approach. The two formulae become identical at small values of the susceptibility but the Onsager theory predicts that the susceptibility will only become infinite at the absolute zero. Pirenne (1949), however, has shown how it may be logically modified to allow the existence of a Curie point.

The results which are frequently used are the modifications which these theories make to the susceptibility. Denoting by χ_0 the susceptibility calculated with allowance for the effects of the crystalline electric field alone, the following expressions are obtained for the susceptibility with dipole interaction:

$$\text{Lorentz } \chi=\chi_0\left(1-\frac{4\pi\chi_0}{3}\right)^{-1} \qquad \ldots\ldots(7a)$$

$$\text{Onsager } \chi=\chi_0\left[\frac{3}{4}-\frac{3}{16\pi\chi_0}+\left(\frac{9}{16}+\frac{3}{32\pi\chi_0}+\frac{9}{256\pi^2\chi_0^2}\right)^{\frac{1}{2}}\right] \ldots\ldots(7b)$$

Both of these approaches are phenomenological, in contrast to the third which is the quantum mechanical method developed by Waller (1936) and in particular by Van Vleck (1937). Van Vleck takes as the Hamiltonian of the system,

$$\mathcal{H}=\sum_i \Phi_i+\sum_{j>i}\sum W_{ij}-H_e\sum_i(\mu_i)_z \qquad \ldots\ldots(8)$$

where μ_i is the magnetic moment of the ith ion, Φ_i the energy due to the crystalline electric field and H_e an external field applied along the z axis. The term W_{ij} represents the interactions between two ions

$$W_{ij}=\frac{1}{r_{ij}^3}\left[\boldsymbol{\mu}_i.\boldsymbol{\mu}_j-\frac{3(\boldsymbol{\mu}_i.\mathbf{r}_{ij})(\boldsymbol{\mu}_j.\mathbf{r}_{ij})}{r_{ij}^2}\right]+V_i \qquad \ldots\ldots(9)$$

where r_{ij} is the distance between ions i and j. The first term is the familiar expression for the potential energy of two dipoles, the second is an exchange term. Van Vleck takes $V_{ij}=v_{ij}(\boldsymbol{\mu}_i.\boldsymbol{\mu}_j)$, where v_{ij} is a constant given in formula (8) of Van Vleck's paper, so that the energy due to exchange is proportional to the cosine of the angle between the individual dipoles. This is referred to as isotropic exchange. The essence of the method is to write down the partition function and to expand it in a power series in $1/T$, obtaining

$$Z=\text{Spur}\left\{1-\frac{\mathcal{H}}{kT}+\frac{\mathcal{H}^2}{2k^2T^2}-\ldots\right\} \qquad \ldots\ldots(10)$$

On account of the theorem concerning the invariance of the spur of a matrix, the above spurs may be evaluated in any representation and it is not necessary to attempt to solve the eigenvalue problem. Herein lies the advantage of the method, which has been much used subsequently in the theory of paramagnetism (see for example Bleaney 1950 a, Cooke, Duffus and Wolf 1953, Daniels, 1953 a, Garrett 1950 b).

Van Vleck's theory gives the following results at high temperatures (i.e. $T \gg \tau$)

$$\frac{S_0 - S}{R} = \frac{Q\tau^2}{12T^2} \qquad \ldots\ldots(11)$$

where $\tau = Ng\beta^2 J(J+1)/k, \qquad S_0 = R \log (2J+1)$

and $$Q = \frac{1}{N^2} \sum_{i} r_{ij}^{-6}(2+v_{ij}^2).$$

Q is a factor which depends upon the type of crystal lattice and which Van Vleck evaluates for simple, body-centred and face-centred cubic structures. He also finds that to this approximation the above entropy reduction and that due to the effect of the crystal field are additive, a fact which is made use of in computing the correction to the ' Brillouin entropy ' for magnetization at 1°K (see, for example, Gardner and Kurti 1954). The expression for the susceptibility, neglecting exchange, is given by

$$\chi = \chi_0(1 - \frac{4\pi}{3}\chi_0 + 12\eta\chi_0^2)^{-1} \qquad \ldots\ldots(7c)$$

where $$\eta = \frac{Q}{12}\left[1 + \frac{3}{8J(J+1)}\right].$$

The effect of isotropic exchange on the high temperature behaviour of χ can be included as a Curie–Weiss Δ, which is simply related to the constant a in the expression for the entropy, $(S_0 - S)/R = a/RT^2$, by the relation $2a/R\Delta^2 = 3/4f$, where $2f$ is the number of nearest neighbours. It was formerly thought that exchange effects in the dilute salts used in this field would be negligible on account of the comparatively large distance between ions. The theory of indirect, or anisotropic, exchange [' superexchange '] however, shows that it is possible for exchange to take place through intervening atoms (Kramers 1934). Opechowski (1948) considered a general type of exchange interaction and showed that in this case also a relation exists between Δ and a. This relation depends on the interaction constants which, unfortunately, are not easily calculated. One important conclusion drawn is that, whereas the contributions to the $1/T^2$ term in the entropy from dipole–dipole interaction and isotropic exchange are additive, this is not so for anisotropic exchange. The theory has been extended further by Daniels (1953 a) to cover the case of ions having anisotropic g-factors.

(d) Hyperfine splitting.

The pattern of the energy levels for a magnetic ion may be further modified if its nucleus possesses a spin. There are two types of interaction which can cause this; the interaction between the nuclear magnetic moment and the magnetic field caused by the electrons in the unfilled shell, and the interaction between the nuclear electric quadrupole moment and the gradient of the electric field at the nucleus. A further source of entropy associated with the $2I+1$ magnetic substates of the nucleus is provided, the overall splitting of which is usually of the order of 0·01 degree.

The theory of hyperfine structure in paramagnetic salts has been given by Abragam and Pryce (1951), and the effect can be described by adding terms to the spin Hamiltonian. For substances where the crystalline electric field has axial symmetry the extra terms are

$$AS_zI_z+B(S_xI_x+S_yI_y)+Q[I_z^2-\tfrac{1}{3}I(I+1)]. \quad \ldots\ldots(12)$$

The constants A and B describe the splitting due to the magnetic interaction and Q describes that due to the quadrupole interaction. These constants may be determined from paramagnetic resonance experiments. The pattern of the energy levels due to these interactions has been discussed in a paper by Bleaney (1951 a) to which we shall again refer in § 8. The effect of the hyperfine splitting is to give rise to a Schottky-type specific-heat anomaly, wherein $C \propto 1/T^2$ at 'high' temperatures. The constant of proportionality has been derived by Bleaney (1950 a) who also showed that at such temperatures the hyperfine splitting has no effect on the susceptibility. Furthermore, the contributions to both entropy and susceptibility from hyperfine splitting and direct interactions between ions are additive as far as the term in $1/T^3$ (Daniels 1953 a).

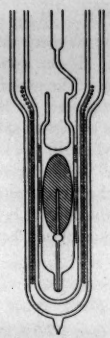
2.2. *Experimental*

The salt specimen, usually spheroidal in shape (see § 3.1), is enclosed in a vacuum space the outer wall of which is in contact with liquid helium. The temperature of the latter is reduced as far as possible by pumping, and temperatures in the range 0·9°–1·3°K are usual.

The magnetic cooling process comprises three stages: (i) The salt is magnetized isothermally, the heat of magnetization being conducted to the liquid helium by 'exchange gas' (helium at a pressure of 10^{-3}–10^{-2} mm Hg) introduced into the vacuum space. (ii) When, after a suitable time interval, the heat transfer is complete, the salt is thermally isolated from its surroundings by pumping off the exchange gas. (iii) The magnet is now de-energized and during this isentropic process the temperature of the salt falls. The specimen thereupon acts as its own 'getter' in removing traces of residual gas by condensation. The success of the adiabatic demagnetization process depends upon the degree to which the salt is thermally isolated from its surroundings. It must be suspended or supported by materials of low thermal conductivity, and room-temperature radiation must be efficiently screened off.

Figures 2 and 3 show two typical cryostats; the essential features are common to both but there are differences in detail. The paramagnetic salt may be in the form of a single crystal, a dense compress of small crystallites, or a loose powder held in a glass or plastic container. By compressing powdered salt under a pressure of a few tons per square inch a specimen is obtained of a density very little less than that of the crystal, and which may, with care, be shaped as desired. The specimen is either suspended by fine nylon fibres in a cylindrical frame which slides into the vacuum case, or is supported rigidly by a thin-walled tube of glass or some other good thermal insulator. Care must be taken to avoid loss of the water-of-crystallization of hydrated salts; for instance the vacuum chamber must never be evacuated unless the specimen is first cooled well below 0°c. Further experimental details relevant to thermal insulation will be given in § 5.

The determination of the susceptibility, which is required for both the study of magnetic properties and the indication and measurement of temperature (see § 3), is usually made by an induction method (see, however, the balance method used by de Haas, Wiersma and Kramers 1933 b). A coil of a few hundred to a few thousand turns of thin copper wire, immersed in liquid helium or liquid hydrogen so that its electrical resistance is small and constant, surrounds the specimen. Its self-



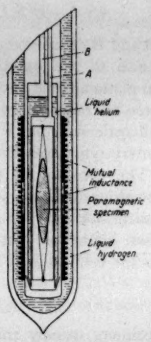


Figure 2. Demagnetization cryostat (*after* Casimir, de Haas and de Klerk 1939 a). An ellipsoidal salt specimen is shown mounted on a glass pillar within a vacuum jacket. This is surrounded by two concentric dewar vessels, the inner one containing liquid helium and the outer one liquid hydrogen (or nitrogen). The secondary coil of the mutual inductance is wound on to the vacuum jacket and the primary on to the ' tail ' of the helium dewar.

Figure 3. Demagnetization cryostat (*after* Kurti 1952). The mutual inductance is wound on to a former which fits over the vacuum case and is surrounded by liquid hydrogen. The liquid helium cryostat is pumped through tube B. Exchange gas may be admitted and removed from the salt space through tube A.

inductance, or more commonly its mutual inductance relative to a second concentric coil, being a linear function of the susceptibility, serves for this measurement. Since the change in mutual inductance due to the change of magnetic susceptibility is small, it is customary to reduce the total circuit mutual inductance by means of compensating coils. The susceptibility may be determined ballistically by observing the deflection of a ballistic galvanometer in series with the secondary coils when the current in the primary coils is reversed. Alternatively, it may be measured by an a.c. method, e.g. a Hartshorn mutual inductance bridge. The frequency is normally within the range 50–500 c/s and a vibration galvanometer or a cathode-ray oscilloscope (in either case preceded by an amplifier) serves as a null detector (Casimir, de Haas and de Klerk 1939 a, Benzie and Cooke 1950 a†).

Compared with the ballistic method, the a.c. method has the advantages of high sensitivity and of being capable of adaptation for automatic recording. On the other hand, at very low temperatures, where relaxation and hysteresis effects become important, the frequency dependence of the susceptibility and the heating effect caused by the alternating primary field are a disadvantage from which the ballistic method does not suffer. The measuring fields are, in both methods, of the order of a few oersteds and may be reduced in favourable cases to a few tenths of an oersted.

The entropy reductions for any given H/T for ideal paramagnetic salts may be obtained from the tabulated values referred to at the end of § 2.1 (*a*) (as mentioned in § 2.1 (*b*), departure from ideal behaviour due to crystalline field effects will

† *Note added in proof.* See also R. A. Erickson, L. D. Roberts and J. W. T. Dabbs, *Rev. Sci. Instrum.*, 1954, **25**, 1178.

necessitate a small correction to the entropy reduction so calculated). In order to enter the region where co-operative effects may be investigated it is essential to reduce the entropy to a value below $R \ln 2$. It is clear that with the limited magnetic fields usually at one's disposal it is of great advantage to reduce the starting temperature as much as possible, since this is equivalent to a proportionate increase of the magnetic field. The temperature that can be reached by pumping on liquid helium is mainly determined by the film of He II creeping up the walls of the container, and is higher than one could achieve if liquid helium were a normal liquid (Rollin 1936, Kurti, Rollin, and Simon, 1936, Rollin and Simon 1939). It has been shown, however, that by suitable design it is possible to reach about 0·8°K even with pumps of relatively small speed, such as 2–3 litres per second (Blaisse, Cooke and Hull 1939, Cooke and Hull 1939, Lasarew and Esselson 1941, Ambler and Kurti 1952 a). Taking 1°K as a round figure for the starting temperature, one may see from the tables that a field of 10–15 kilo-oersteds should bring one into the region of the co-operative anomaly. But if one wants to remove the magnetic entropy to within a few per cent of its full value much higher fields are required, e.g. 4×10^4 oersted deg^{-1} would be necessary to get into a temperature region corresponding to room temperature for the case of ordinary ferromagnetic metals such as iron or nickel.

The simplest way of increasing the field obtainable from a given piece of equipment is to reduce the size of that part of the cryogenic apparatus that has to be brought into the magnetic field. Now to ensure tolerable warm-up rates, the diameter of the salt specimen should not be less than about 5 mm and this would bring to about 25 mm the overall outside diameter of the apparatus, which usually consists of two or three layers of liquefied gases with intervening vacuum spaces. Somewhat larger cryostats are, however, more convenient to assemble, and in most demagnetization apparatus a compromise is reached with outside diameters of 35–50 mm permitting samples of 15–25 mm diameter to be used.

Fields of 10–15 kilo-oersteds in pole gaps of up to 40–50 mm can be obtained from medium-sized laboratory electromagnets equipped with cooled windings, and difficulties arise only when fields of about 20 kilo-oersteds are required. There are two ways of producing fields of this intensity and extension: either with electromagnets having large diameter cores and tapered pole pieces (e.g. the electromagnets of the Kamerlingh Onnes Laboratory in Leiden and of the Bellevue Laboratory near Paris) or with efficiently cooled high-powered iron-free solenoids. This latter method has been considerably developed in recent years, due mainly to the work of Cockcroft (1928), Bitter (1936), Ashmead (1944) and Tsai (1947), and the technique of winding such coils has received much attention (Gaume 1946, Daniels 1950, Daniels 1953 b). Solenoids consuming 1000 kw or more and producing fields of 40–50 kilo-oersteds for considerable periods of time have been in operation in the last few years. The following rough and ready formula (Hudson 1949 a) is useful for obtaining a conservative estimate of the field produced by an iron-free solenoid: $H(\text{kilo-oersted}) = 4(W/D)^{1/2}$. Here W is the power consumed in kilowatts, D the internal diameter of the solenoid in centimetres and it is assumed that the shape of the coil is not far off optimum, i.e. that both the outside diameter and the length are between $3D$ and $5D$. The temperature rise of the copper conductor has been taken into account and a space factor consistent with efficient cooling has been allowed for.

Solenoids with liquid air or liquid nitrogen cooling have also been used (Kurti 1933, Collins 1950). They have the advantage of lightness, simple construction,

and a reduction of the necessary power by about a factor of six, corresponding to the reduced electrical resistivity. If the field need not be kept on for long periods the consumption of liquid nitrogen (about 1 litre/min for 10 kilo-oersteds in a coil of 50 mm inside diameter) is not prohibitive.

The relative merits of iron magnets and iron-free solenoids from the point of view of magnetic cooling experiments (leaving aside questions of capital cost, running cost, etc.) may be summarized as follows. Iron magnets have the advantage that, by virtue of their relatively high self-inductance in their pre-saturation range, they act as chokes and reduce the ripple which is a troublesome feature of d.c. supplies provided by rectifiers or motor generators; with solenoids, separate chokes or smoothing circuits may have to be installed. As to ease of handling, small iron magnets (say up to two tons) can be almost as easily moved from the cryostat as solenoids; with large iron magnets one has to move the cryostat (de Haas and Wiersma 1933, Kurti, Lainé, Rollin and Simon 1936 b, de Klerk and Hudson 1954).

The main advantage of working with an iron-free solenoid lies in the wider possibilities it offers for measuring the magnetization. With the inductance method it is relatively easy to determine the differential susceptibility in an external field, and if measuring field and external field are parallel to each other the total magnetic moment can be obtained by integration. It was mainly for this reason that Giauque and his collaborators did all their experiments with an iron-free solenoid. With an iron magnet the magnetic moment can of course be measured directly by means of a balance method but because of its other shortcomings this technique has not been used since the first experiments of de Haas and Wiersma.

The difficulty of measuring the susceptibility by an inductance method in the presence of an iron magnet lies in the influence of the iron on the self or mutual inductance, especially as this effect varies with the degree of saturation of the pole-pieces. Casimir and de Haas (1940), however, succeeded in determining the magnetization by means of a mutual inductance set at right angles to the external field produced by an iron-magnet. If I and H_e are the magnetization and external field respectively, and χ the differential susceptibility measured at right angles, we have for an isotropic, spherical specimen $I=\chi H_e$. For ellipsoidal specimens the anisotropy of the demagnetizing factor necessitates a correction.

Many important experimental details of magnetic cooling technique have been omitted from this section; they will be mentioned in later sections under the appropriate topics. A detailed description of the technique practised in Oxford has been given by Hull (1947), and of a recent magnetic cooling installation at the U.S. National Bureau of Standards by de Klerk and Hudson (1954).

§ 3. The Determination of Temperatures below 1°K

3.1. *The 'Magnetic Scale'*

The determination of temperatures below 1°K presents unusual difficulties in that all the conventional methods fail. For example, gas thermometry is inapplicable owing to the rapid decrease of the vapour pressure of helium, which by 0·5°K is of the order of 10^{-5} mm Hg. The problem is to find a suitable thermometric parameter, i.e. any easily measured physical property which is a single-valued and sufficiently rapidly varying function of the absolute temperature. The magnetic susceptibility χ of the actual salt used for the cooling satisfies this condition over an appreciable temperature range and is most commonly used. Since in the helium

range χ obeys Curie's law, i.e. is inversely proportional to the absolute temperature T, it is logical to define a 'Curie' (or magnetic) scale of temperature T^* by extrapolating Curie's law into the demagnetization range.

A magnetic temperature is defined by using the value of the 'external' susceptibility, i.e. $\chi_e = I/H_e = c/T^*$, where c is the Curie constant per cm³. On account of demagnetizing effects we obtain different scales for differently shaped specimens, and it was proposed by Kurti and Simon (1938 b) to reduce them all to that appropriate to a spherically shaped specimen, T^*, by means of the relation $T^* = T^* + \beta$ where $\beta = (n_s - n)c$, n_s being the demagnetizing coefficient for a sphere ($=4\pi/3$) and n that for the specimen actually used. It is essential that specimens should always be ellipsoidal in shape so that n can be calculated. The proposal to correct the various magnetic scales to the particular one for a sphere was not made arbitrarily but has a special advantage.

One may define an 'ideal susceptibility', χ_0, as the ratio of the magnetic moment to the field actually *acting upon the ions*, $\chi_0 = M/H_{loc}$, and this quantity should obey Curie's law to a lower temperature than does χ_e. In the higher temperature region, one may accurately describe the effect of dipole interactions by the Lorentz internal field; for a sphere, the Lorentz and demagnetizing fields cancel and $H_{loc} = H_e$, $\chi_0 = \chi_e$, and thus T^* should be a better approximation to T than any other magnetic temperature.

In fact, in the absence of appreciable Stark splitting T^* should be equal to the absolute temperature in the range of validity of the Lorentz treatment. (Some salts actually obey a Curie–Weiss law, $\chi = c/(T - \Delta)$ and Δ may be determined by careful calibration in the liquid helium range or from measurements of the field dependence of the adiabatic susceptibility (Garrett 1950 a).)

For the purpose of correlating results obtained for differently shaped specimens the proposal of Kurti and Simon remains of use, of course, even outside the range where $T^* = T$. They pointed out, however, that at the very lowest temperatures where co-operative effects occur, the values of T^* computed from the values of T^* obtained with differently shaped specimens would not necessarily agree, and hence in quoting results the value of β used should always be given. The experiments of Steenland, de Klerk, Potters and Gorter (1951) showed this to be the case.

3.2. *Calorimetric Methods*

From the Kelvin definition of thermodynamic temperature, $T = dQ/dS$, it follows that absolute temperatures may be obtained directly if the change in entropy corresponding to the introduction of a known amount of heat can be determined. Using the magnetic temperature T^* as a thermometric parameter one may write $T = (dQ/dT^*)/(dS/dT^*)$. The (S, T^*) relation is obtained by measuring T^* immediately after each of many demagnetizations, the corresponding entropies being calculated from the initial conditions of H and T. (See § 2.1 (*a*).) It is the determination of the (Q, T^*) relation which presents the greatest problem. Ideally, one simply introduces a uniformly distributed known amount of heat dQ and measures the resulting change in T^*, i.e. the experiments comprise, essentially, measurement of a specific heat C^*, defined by

$$C^* = dQ/dT^* = T(dS/dT^*). \qquad \ldots\ldots(13)$$

In practice it is difficult to supply heat uniformly to the sample and, especially, to make adequate correction for the effects of stray heating. The latter is always a

surface effect and at very low temperatures, where the thermal diffusivity becomes very small, a pronounced temperature gradient will persist across the sample (Kurti and Simon 1938 c, Cooke and Hull 1942). The measured susceptibility is then a mean value, so that the derived value of T^* is not unique and the resulting (Q, T^*) curve cannot be correlated with the known (S, T^*) relation. One ingenious method of dealing with this problem has been discussed by Daniels and Kurti (1954). It relies on the fact that in powder specimens at very low temperatures the thermal diffusivity is so extremely small that, to a good approximation, the state of affairs may be described by the following simple model. There are two distinct zones (i) a core which, under the influence of background heating alone, remains at constant temperature, and (ii) a surrounding shell at a much higher, but also constant, temperature. As the stray heat influx continues the boundary between the zones moves steadily inwards, and if the further assumption is made that the deliberately introduced heat has a noticeable effect on the core alone, the analysis of the observations is tractable.

As far as deliberately introducing heat into the salt specimen is concerned the most obvious method is that of passing current through a resistance which is in good contact with the salt. The method fails below a few tenths of a degree Kelvin through deterioration in the quality of this contact. An elegant method introduced by Kurti and Simon (Simon 1935, Kurti and Simon 1935 c) is to irradiate the specimen with gamma rays. A good degree of uniformity may be achieved by symmetrically disposing a number of sources at a suitable distance, or by automatically rotating two sources, set equidistant from the specimen and diametrically opposite each other. If the absorption remains independent of temperature the rate of heating will remain constant, and the source may be calibrated by continuing the measurements into the 'high temperature' region where the relation between T^* and T is known already (see §§ 3.1 and 3.4).

At very low temperatures, in the region of the susceptibility maximum, the a.c. susceptibility becomes complex (Giauque and MacDougall 1935 b) thus one may write

$$\chi=\chi'-i\chi'' \qquad \ldots\ldots(14)$$

and it may be shown that the application of an alternating magnetic field dissipates energy in the salt at a rate given by $dW/dt=\frac{1}{2}h_0^2\omega\chi''$ where h_0 is the peak-value of the a.c. field, ω is the angular frequency and χ'' the loss-component of the susceptibility (Casimir, de Haas and de Klerk 1939 b). For zero stray heat influx, we have $dQ=dU-dW=0$ hence $C^*=dU/dT^*=dW/dT^*$. While χ'' is increasing with decreasing entropy the a.c. heating tends to restore homogeneity of temperature, upset by stray heating effects, and the degree of success of this method depends upon the degree to which this compensation is provided. (If χ'' also shows a maximum at a certain entropy the method cannot be used at lower entropies, since inhomogeneities would then be magnified.)

Yet another method of doing work upon the substance is by executing a large number of hysteresis cycles during ballistic measurements. The area of a hysteresis loop is a measure of the energy dissipation per cycle (Steenland, de Klerk and Gorter 1949).

In practically all cases where a given substance has been investigated in the same entropy range by the two calorimetric methods, gamma-ray and χ'' heating (this has never been done, unfortunately, in any one laboratory), the χ'' method has given much lower temperature values. It has been suggested (de Klerk, Steenland

and Gorter 1948, 1949) that the discrepancies may be due to unreliability of a gamma-ray source as a constant-rate heat supply when applied to a paramagnetic salt at very low temperatures. To particularize, the absorbed energy might be, in part, stored elsewhere in the crystal rather than all being communicated at once as heat to the spin system. This question has been discussed at some length by Platzman (1953) who concludes that a partial inhibition of the final conversion of the absolute energy to heat is entirely possible, especially through ' permanent ' dissociation of molecules. Kurti and Simon (1953) have commented on Platzman's analysis, pointing out that the experimental evidence is against the possibility of such effects being serious for the temperature range and substances investigated so far. Finally, it should be mentioned that the loss-component χ'' and the remanence may also be used as thermometric parameters (de Klerk, Steenland and Gorter 1949) as well as for introducing heat. In the final analysis, the accuracy of all types of low-temperature calorimetric determination depends upon the extent to which the stray heat influx may be reduced by improvement in design of apparatus (see § 5.1).

3.3. *Magnetic Methods*

The temperature change upon isentropic application of a magnetic field can be calculated from equation (1), and for a finite change,

$$[\Delta T]_S = -\int_H^{H'} \left(\frac{\partial M}{\partial S}\right)_H dH. \qquad \ldots\ldots,(15)$$

By obtaining a lattice of (M, H) curves at various entropies ranging between ' high ' known temperatures and low unknown temperatures, the latter may be determined from this relation. This is essentially the method proposed by Giauque (Giauque 1936, 1953, Giauque and MacDougall 1938). Since the value of the absolute temperature being determined is very much less than 1°K, and the high known temperature is about 1°K, the former enters as a small difference between two relatively large quantities. Furthermore the derivation of ΔT can be seen from equation (1) to be a rather involved procedure, so that extremely accurate magnetic measurements are necessary to achieve a moderate degree of accuracy in the absolute temperature. Since at least comparable accuracy can be obtained by simpler methods, the above procedure has not been used extensively.

It is an important point, however, that this method is available and that absolute temperature determinations may be made without the necessity of deliberately introducing heat into the salt specimen. The method has recently been discussed at length (Giauque 1953) and its merits *vis-à-vis* the calorimetric procedure outlined.

The above type of experiment permits the determination of absolute temperatures in both ' zero ' and finite magnetic field. A second approach to the latter was described by de Haas and Wiersma (1936) which makes use of the fact that the magnetic moment M during adiabatic demagnetization remains constant down to a very small value of the magnetic field. Using a thermodynamic argument they obtained the general relation for this situation, that M is a function of $(H-K)/T$, where K depends only on the initial conditions of H and T. The constant K can, in principle, be determined by accurate measurements of magnetization curves in the liquid helium region. (Note that for an ideal paramagnetic K is zero.) The values obtained for very low temperatures are highly sensitive to the value of K, thus here again extremely accurate magnetic measurements are required to give a moderate accuracy for the values of the absolute temperature.

3.4. *The ' Theoretical Method '*

This is an indirect method of obtaining a (T^*, T) correlation based on the Hebb–Purcell theory, and has proved to have a limited applicability. Paramagnetic resonance experiments have shown that, in many cases, the Hebb–Purcell theoretical treatment contains over-simplified assumptions concerning the symmetry of the crystalline electric field and neglects hyperfine structure and exchange effects which are often significant. It is, however, a very simple method and the experimental determinations involved can be made with high accuracy. In no case does the range of validity extend very far below 0·1°K.

The theory [see §§ 2.1 (*b*), (*c*)] leads to expressions for the susceptibility and the entropy as functions of the absolute temperature. The principle of the method is to combine these two relations to obtain a third, or (χ_0, S), relation. This may be checked by direct experiment and within the range of agreement it is assumed that the two original formulae are individually correct, whence T may be calculated from the experimentally observed χ_e if the relation between χ_e and χ_0 is known. The latter is an ' idealized susceptibility ' expressed in terms of the field actually acting upon the ions, H_{loc}, and differs from χ_e on account of the effects of magnetic interaction and the external shape of the specimen. The latter is taken care of by reducing the measured susceptibility to the case of a sphere and the former may be taken into account by theories developed by Lorentz, Onsager and Van Vleck. In general, the Onsager approach is found to be most satisfactory.

The ' Leiden convention ' is to compare theory and experiment in terms of magnetic temperature and to reverse the above procedure by correcting experimentally observed values of T^* to those which would be obtained without interactions. Use of the three methods for allowing for the interactions leads to symbols T^*_{Lor}, T^*_{Ons} and T^*_{VV} [T^*_{Lor} is identical with T^*].

3.5. *Secondary Thermometers*

Carbon resistors offer many practical advantages such as compactness, stability under the heating which is sometimes necessary during assembly of apparatus, and essentially field-independent behaviour. Colloidal carbon films were investigated by Giauque, Stout and Clark (1938) and some data for carbon-composition resistors down to 0·3°K have been published by Clement, Quinnell, Steele, Hein and Dolecek (1953). The latter authors conclude that carbon-composition resistors exhibit more reproducible behaviour after warming to room temperature and being recooled, while the colloidal carbon films offer a much wider range of thermometric sensitivity. Howling, Darnell and Mendoza (1954) have investigated a carbon resistor down to 0·05°K, using the resistance as a thermometric parameter (rather than the susceptibility of the paramagnetic salt) in making absolute temperature determinations. (In several recent experiments (see § 6) carbon resistors have been used to indicate temperatures below 1°K, after calibration against a ' magnetic thermometer ', but no pertinent data have been published.)

Leaded phosphor bronze wires have been studied as resistance thermometers down to $T^*=0{\cdot}027°$ (ferric ammonium alum) by Allen and Shire (1937) and by Van Dijk, Keesom and Steller (1938) who found a linear dependence of R on T^* down to 0·25°. The resistance is markedly dependent on current and on magnetic field by virtue of the presence of small occlusions of superconducting lead, which are, in fact, responsible for the temperature-dependent resistivity.

3.6. *The Results*

Absolute temperature determinations below 1°K have been made on relatively few paramagnetic salts although one or two of these have received special attention from various investigators. Salts for which data have been obtained will be dealt with below in the (approximately) chronological order of their first investigation. The brief summaries of spectroscopic data have been taken largely from the review article of Bleaney and Stevens (1953).

(*a*) *Ferric ammonium alum*, $Fe_2(SO_4)_3(NH_4)_2SO_4.24H_2O$.

Gramme ionic weight 482·2. Density† 1·71 g cm^{-3}.

The free Fe^{3+} ion is in a $^6S_{5/2}$ state and a crystalline electric field of cubic symmetry would split the six-fold degenerate ground level into a doublet and a quadruplet. A field of lower symmetry would, however, give rise to a separation into three doublets (Meijer 1951), as appears to be required by the shape of the 'T^{-2} region' of the specific heat curve (Benzie and Cooke 1950 b). Resonance measurements on diluted specimens showed g to be isotropic and equal to 2 within less than 1% (Ubbink, Poulis and Gorter 1951, Whitmer and Weidner 1951).

The relation between T^* and T was first investigated, using gamma-ray heating, by Kurti and Simon (1935 c) and has been published in graphical form by Cooke (1949); the data below 0·2°K are given in table 1.

Table 1. (T, T^*) Correlation for Ferric Ammonium Alum

T(°K)	T^*(°K)	T(°K)	T^*(°K)	T(°K)	T^*(°K)
0·015	0·0720	0·043	0·0850	0·090	0·1390
0·020	0·0702	0·045	0·0910	0·100	0·149
0·025	0·0683	0·050	0·0970	0·120	0·168
0·030	0·0670	0·055	0·1008	0·140	0·183
0·035	0·0666	0·060	0·1050	0·160	0·198
0·040	0·0666	0·070	0·1150	0·180	0·214
0·042	0·0668	0·080	0·1273	0·200	0·229

The Curie point was found to be 0·042°K, reached from initial conditions of H_i/T_i=11500. Steenland, de Klerk, Potters and Gorter (1951) find, in general, lower values of absolute temperature and give 0·030°K for the Curie point; they introduced energy into their specimen by utilizing a.c. losses and, below the Curie point, d.c. hysteresis losses.

De Klerk (1948) used the 'theoretical method' (§ 3.4) and obtained agreement with the Hebb–Purcell theory down to 0·27°K using a value of 0·183 degree for the overall splitting, and assuming the doublet–quadruplet pattern. (The paramagnetic relaxation experiments of Benzie and Cooke (1950 b) indicated 0·23 degree for a system of three equispaced doublets; their heat capacity values would correspond with a splitting of 0·20 degree, were the doublet–quadruplet pattern correct.)

(*b*) *Gadolinium phosphomolybdate*, $GdPO_4(MoO_3)_{12}.30H_2O$.

Gramme ionic weight 2520.

The free Gd^{3+} ion is in an $^8S_{7/2}$ state. The eight-fold degenerate ground level is split by a cubic field into two doublets and a quadruplet.‡ The latter lies between

† The densities given here and in subsequent sections refer to room temperature.

‡ See footnote to § 2.1 (*b*)

the two former and the spacing is in the ratio 3:5 (cf. the case of the nitrobenzene sulphonate, § 2.1 (*b*)).

Giauque and MacDougall (1935 a, 1938) measured the specific heat by an induction method of heating and compared their results with the (S, T^*) data. They concluded that, down to 0·15°K, T^* does not differ from T by more than 0·03°K.

(*c*) *Chromic potassium alum*, $Cr_2(SO_4)_3.K_2SO_4.24H_2O$.

Gramme ionic weight 499. Density 1·83 g cm^{-3}.

The free Cr^{3+} ion is in a $^4F_{3/2}$ state but the effect of the crystalline electric field is so strong that the phenomenon of ' quenching of the orbital momentum ' occurs and the ion is in, effectively, a ^4S state. The spin quadruplet remains degenerate in a cubic field but is split into two Kramers doublets by any field of lower symmetry. The magnitude of the splitting, in ' degrees ', has been found to be 0·263 (Casimir, de Klerk and Polder 1940), 0·27 (de Klerk 1948), 0·251 (de Klerk, Steenland and Gorter 1949), 0·245 (Bleaney 1950 b), 0·258 (P. H. Keesom 1948) and $0{\cdot}25_0$ (Ambler and Hudson 1954 a).

The variable behaviour of this salt, as evidenced by the splittings quoted above, is probably due to the occurrence of a structural transition which may be observed at about 80°K (Kraus and Nutting 1941). The individual features of its behaviour appear to be irreconcilable and it is, strictly speaking, not permissible to apply the Hebb–Purcell theory to this salt owing to the as yet unexplained appearance of *two* splittings (Bleaney 1950 c). Any calorimetric type of investigation will yield a root mean square value for the splitting (Daniels and Kurti 1954).

Bleaney (1950 b), using gamma-ray heating, measured absolute temperatures down to 0·05°K and these are given in table 2. Temperatures in the region of the

Table 2. (T, T^*) Correlation for Chromic Potassium Alum

ln (4−S/R)	T(°K)	T^*(°K)	ln (4−S/R)	T(°K)	T^*(°K)
0·008	1·000	1·000	0·187	0·180	0·195
0·022	0·600	0·604	0·226	0·160	0·174
0·034	0·480	0·485	0·271	0·140	0·156
0·047	0·400	0·406	0·329	0·120	0·138
0·057	0·360	0·368	0·404	0·100	0·121
0·072	0·320	0·330	0·503	0·080	0·103
0·091	0·280	0·291	0·616	0·060	0·086
0·120	0·240	0·252	0·682	0·050	0·079
0·158	0·200	0·215	0·720	0·045	0·075

Curie point have been measured by de Klerk, Steenland and Gorter (1949) and by Daniels and Kurti (1954). These are presented as an entropy–temperature curve in figure 4, wherein a marked disagreement between the two sets of measurements may be observed. As usual, at the lowest entropies the Leiden values of T are much smaller than the Oxford values. The triangles in this diagram are points obtained by Ambler and Hudson (1954 a) using the Leiden method of a.c. heating and are of interest in that the temperatures they represent are unusually *high*. The temperatures of the Curie point are 0·004°K (Leiden), 0·011°K (Oxford), and $(H/T)_c = 16000$.

(*d*) *Manganous ammonium sulphate*, $MnSO_4.(NH_4)_2SO_4.6H_2O$.

Gramme ionic weight 391. Density 1·83 g cm^{-3}.

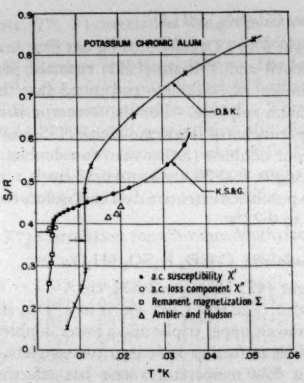


Figure 4. Entropy versus temperature for potassium chromic alum in the region of the Curie point. Curve D. & K., results of Daniels and Kurti (1953) using gamma-ray heating and the ballistic susceptibility as a thermometric parameter. Curve K.S. & G., results of de Klerk, Steenland and Gorter (1949) using heating by a.c. losses and, as thermometric parameter, a.c. susceptibility χ', a.c. loss component, χ'', remanent magnetization Σ. Results of Ambler and Hudson (1954 a) using χ'' heating and χ' as thermometric parameter.

This is a Tutton salt, the crystalline electric field having tetragonal symmetry. The free Mn^{2+} ion is in a $^6S_{5/2}$ state and the interaction with the crystalline electric field is very small. Resonance experiments (Bleaney and Ingram 1951 a) show that $g=2{\cdot}000$ and is isotropic to about 1 part in 10^4.

Adiabatic demagnetization experiments by Cooke and Hull (Cooke 1949) indicated that $T_c=0{\cdot}14°K$ and $(H/T)_c=6200$. Their (T^*, T) correlation was presented in graphical form and is summarized in table 3. Their calibration of the γ-ray source was made, however, by assuming that the Hebb–Purcell theory was valid down to $0{\cdot}5°K$ and that the quadruplet–doublet splitting pattern was correct. Actually, resonance experiments show that the pattern consists of three unequally spaced doublets, which is considerably modified by a surprisingly large hyperfine splitting (Bleaney and Ingram 1951 a, Abragam and Pryce 1951). Steenland, van der Marel, de Klerk and Gorter (1949) made absolute temperature measurements, using χ'' heating, in the region of, and below, T_c. The latter was found to be $0{\cdot}10°K$, which is appreciably lower than the value of Cooke and Hull.

Table 3. (T, T^*) Correlation for Manganous Ammonium Sulphate

T(°K)	T^*(°K)	T(°K)	T^*(°K)
0·13	0·115†	0·315	0·35
0·13	0·15	0·37	0·40
0·16	0·20	0·425	0·45
0·205	0·25	0·48	0·50
0·26	0·30		

† minimum

(*e*) *Gadolinium sulphate*, $Gd_2(SO_4)_3.8H_2O$.

Gramme ionic weight 373. Density 3·010 g cm^{-3}.

Bleaney, Elliott, Scovil and Trenam (1951) reported paramagnetic resonance experiments on gadolinium ethylsulphate and noted that the spectrum is of the same nature in gadolinium sulphate. For the former, g is isotropic and equal to 1·993. The recent experiments of Bogle and Heine (1954) have shown the splitting pattern to consist of four doublets (rather than two doublets and a quadruplet as assumed heretofore) lying at 0, 0·20, 0·48 and 0·82 cm^{-1}.

Van Dijk (1946) has published extensive data on absolute temperatures, measured calorimetrically, down to 0·22°K.

(*f*) *Copper potassium sulphate*, $CuSO_4.K_2SO_4.6H_2O$.

Gramme ionic weight 442. Density 2·22 g cm^{-3}.

This is a Tutton salt. The free Cu^{2+} ion is in a $^2D_{5/2}$ state. In a cubic field the orbital levels split into an upper triplet and a lower doublet. In a tetragonal or rhombic field the doublet is further split into two singlets. The latter are still spin-degenerate and at low temperatures one has effectively $S=\frac{1}{2}$. Bleaney, Bowers and Ingram (1951) find that $g_\parallel=2{\cdot}45$, $g_\perp=2{\cdot}14$. The susceptibility follows a Curie–Weiss law with $\Delta=+0{\cdot}034°$ (Garrett 1950 a, Benzie and Cooke 1950 b) and Garrett (1950 a) finds that $T=0{\cdot}987\ T^*+0{\cdot}033°$ down to $T^*=0{\cdot}025°$.

De Klerk (1946) obtained a (T^*, T) correlation for a powder specimen, T being measured by means of a potassium chromic alum 'thermometer', and reported a much larger value for Δ, viz. +0·052 degree.

Demagnetization experiments by Steenland, de Klerk, Beun and Gorter (1951), in which g was taken to be 2·18, yielded $T_c=0{\cdot}05$°K for $H_i/T_i=11500$ and approximate absolute temperature values in the range 0·0046°–0·025°K. Their lowest values have been criticized, however, by Daniels (unpublished) as being incompatible with the entropy due to hyperfine structure. The latter contributes notably to the specific heat (Garrett 1950 b).

(*g*) *Cobalt ammonium sulphate*, $CoSO_4.(NH_4)_2SO_4.6H_2O$.

Gramme ionic weight 395·2. Density 1·902 g cm^{-3}.

The free Co^{2+} ion is in a $^4F_{9/2}$ state. In a cubic field the seven-fold orbital state is split into two triplets and a singlet, with one triplet the lowest. The latter, with its four-fold spin-degeneracy, is then split into a number of doublets by the combined effect of the spin–orbit coupling and the tetragonal field of this Tutton salt. For the lowest doublet $g_\parallel=6{\cdot}45$, $g_\perp=3{\cdot}05$ (Bleaney and Ingram 1951 b). Garrett (1951 a), taking $g_\parallel=6{\cdot}2$, $g_\perp=3{\cdot}0$ (from early data), obtained (T^*, T) data for the K_1, K_2 and K_3 axes; $T_c=0{\cdot}084$°K.

(*h*) *Chromic methylammonium alum*, $Cr_2(SO_4)_3.(CH_3NH_3)_2SO_4.24H_2O$.

Gramme ionic weight 492. Density 1·645 g cm^{-3}.

The paramagnetic resonance measurements of Bleaney (1950 c) showed there to be a unique value of the ground-state splitting δ. It should therefore be more suitable as a standard thermometric substance than the potassium alum.

Adiabatic demagnetization experiments by Hudson and McLane (1954) indicated that the Hebb–Purcell theory provides a good description of the behaviour down to about 0·08°K, using the Onsager internal field formula and $\delta/k=0{\cdot}269$

degree; the appropriate (T^*, T) correlation was given in a table. Their value of δ/k is in good agreement with the value 0·275 for a powder specimen (de Klerk and Hudson 1953) and 0·27 (Gardner and Kurti 1954). The latter authors measured absolute temperatures by the gamma-ray method of heating in the range 0·013–0·396°K and their results are summarized in table 4. They found $T_c=0{\cdot}02$°K at $(S/R)_c=0{\cdot}500$, $(H/T)_c=13\,900$, and suggest that the difference between their value of δ/k and the paramagnetic resonance value of 0·245 obtained by Bleaney (1950c) indicates that a small amount of anisotropic exchange is contributing to the specific heat.†

Table 4. (T, T^*) Correlation for Chromic Methylammonium Alum

S/R	T^*(°K)	T(°K)	S/R	T^*(°K)	T(°K)
0·350	0·053	0·013	0·900	0·121	0·096
0·400	0·052	0·016	0·950	0·132	0·109
0·450	0·051	0·018	1·000	0·143	0·122
0·500	0·051	0·020	1·050	0·157	0·137
0·550	0·053	0·023	1·100	0·175	0·153
0·600	0·061	0·028	1·150	0·196	0·180
0·650	0·071	0·032	1·200	0·224	0·210
0·700	0·083	0·039	1·250	0·266	0·254
0·750	0·092	0·048	1·300	0·336	0·326
0·800	0·102	0·065	1·350	0·404	0·396
0·850	0·111	0·077			

(*i*) *Cerium magnesium nitrate*, $2Ce(NO_3)_3.3Mg(NO_3)_2.24H_2O$.

Gramme ionic weight 765.

The free Ce^{3+} ion is in a $^2F_{5/2}$ state; the next lowest level to the ground state, $^2F_{7/2}$, is about 2500 cm^{-1} above. Cooke, Duffus and Wolf (1953) assumed the ground state to be split into doublets characterized to a first approximation by $J_z=\pm1/2,\ \pm3/2,\ \pm5/2$. The lowest doublet was identified as $J_z=\pm1/2$ with $g_\perp=1{\cdot}84$, $g_\parallel=0{\cdot}25\pm0{\cdot}05$.

Daniels and Robinson (1953) employed gamma-ray heating in demagnetization experiments and found that $T^*=T$ down to 0·006°K. Their (T^*, T) data between $T=0{\cdot}003°$ and 0·006°K are summarized in table 5.

Table 5. (T, T^*) Correlation for Cerium Magnesium Nitrate

S/R	$T(10^{-3}$ deg K)	$T^*(10^{-3}$ deg K)	S/R	$T(10^{-3}$ deg K)	$T^*(10^{-3}$ deg K)
0·100	3·08	3·20	0·525	3·76	4·35
0·300	3·08	3·20	0·550	4·25	4·72
0·350	3·08	3·28	0·560	4·50	4·90
0·400	3·08	3·38	0·570	4·79	5·09
0·425	3·08	3·49	0·580	5·12	5·32
0·450	3·12	3·66	0·590	5·46	5·56
0·475	3·23	3·88	0·600	5·86	5·86
0·500	3·43	4·09			

In this salt, owing to the smallness of the magnetic specific heat, the contribution to the specific heat from the lattice is not negligible and has been estimated to become comparable with the magnetic specific heat at about 0·5°K. It should be

† The most recent analysis of paramagnetic resonance data (J. M. Baker, unpublished), assuming a small rhombic component in the crystal field, indicates that $\delta/k=0{\cdot}255$. The possible existence of some anisotropic exchange still remains.

stressed that the property of obeying Curie's law down to 0·006°K potentially makes this salt an extremely valuable ' thermometer '. Owing to its marked anisotropy it cannot be used in powdered form; on the other hand, this feature may be of considerable use in another connection (see § 7.4).

3.7. *Note on Discrepancies*

Although a rather large amount of data on temperature measurements has now been published it is largely confined to a few salts and to the region above 0·1°K. Below this point the data are sparse and the work of one laboratory rarely overlaps that of another; where this does occur the agreement is generally bad. For example, the Curie point of ferric ammonium alum is given as 0·030°K (Leiden) and 0·042°K (Oxford), and that of chromic potassium alum as 0·004°K (Leiden) and 0·011°K (Oxford). The questioned validity and the defence of the gamma-ray method have already been mentioned in § 3.2; it appears most probable that the cause of the discrepancies lies elsewhere (see below). It is now generally agreed that in order to resolve the present difficulties, measurements should be made using both methods on the same specimen in several laboratories.

The two curves of figure 4 (§ 3.6 (*c*)) illustrate the magnitude of the discrepancy between the Leiden and Oxford temperature values. Since the triangle points were obtained using the ' Leiden method ' (Ambler and Hudson 1954 a), these points provide evidence, though slight, that the divergence is not due to invalidity of the gamma-ray method. The most probable cause is the difficulty of making an accurate correction for the effects of stray heating (see § 3.2) and a general improvement in experimental technique to permit measurements under conditions of negligible ' drift ' is now highly desirable.

§ 4. Properties of Paramagnetic Salts below 1°K

§ 4.1. *Susceptibility*

The susceptibility may be measured either ballistically or by an a.c. bridge method and, as mentioned before, is readily obtained as a function of entropy. As the entropy decreases it is observed that the ' zero-field ' susceptibility first increases at a rate roughly proportional to the square root of the entropy diminution, then much more rapidly and ultimately goes through a maximum. In the region of the maximum the d.c. susceptibility is usually found to be a few per cent higher than the a.c. susceptibility; (χ, S) curves for chromic potassium alum and chromic methylammonium alum are shown in figures 5 (*a*) and 5 (*b*) respectively. One may note that in the methylammonium alum the rapid increase in χ sets in at entropy $R \ln 2$, in accordance with expectations, whereas this does not occur in the potassium alum until a much lower entropy. This latter behaviour is not understood.

From the theoretical point of view the relationship between χ and T is, of course, more useful than that between χ and S. Since χ is inversely proportional to T^* this is equivalent to a (T^*, T) correlation, discussed previously.

4.2. *Relaxation and Hysteresis Phenomena*

The decrease of entropy below the value $R \ln 2$ and the appearance of a susceptibility maximum are evidence of the temperature decreasing to such a point that kT has become of the order of the magnetic interactions. The latter give rise to a co-operative effect with a final spatial ordering of the spins towards an entropy

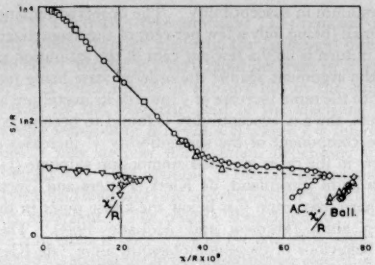


Figure 5(*a*). Chromic potassium alum: entropy plotted against susceptibility (*after* de Klerk 1952, p. 215, fig. 44.3).

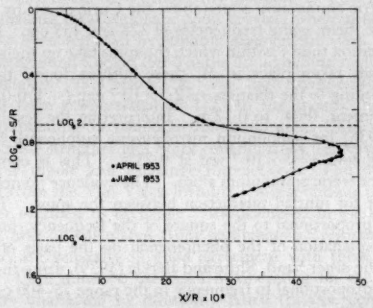


Figure 5(*b*). Chromic methylammonium alum: entropy plotted against a.c. susceptibility (*after* Hudson and McLane 1954).

of zero at 0°K. Experiments suggest that the most probable arrangement is an antiparallel, or antiferromagnetic, one.

In the early experiments of Kurti, Lainé, Rollin and Simon (1936 a) upon ferric ammonium alum an associated phenomenon was observed, namely the appearance of hysteresis effects below the entropy corresponding to χ_{max}. The remanence is found to increase at first with increasing magnitude of the cycled field, but the hysteresis disappears in fields of about 100 oersteds and greater. In the case of chromic methylammonium alum, the remanence when plotted as a function of entropy goes through a maximum (Hudson and McLane 1954). This type of behaviour has not been reported for the other salts investigated, e.g. chromic potassium alum (Steenland, de Klerk and Gorter 1949, Ambler and Hudson 1954 a), manganous ammonium sulphate (Steenland, van der Marel, de Klerk and Gorter 1949), ferric ammonium alum (Kurti, Lainé, Rollin and Simon 1936 a, Steenland, de Klerk, Potters and Gorter 1951). In these latter cases the remanence is found to increase continuously as the entropy is lowered.

Conventionally the ' Curie point ' is taken to be defined by the appearance of remanence, which generally occurs at a slightly higher entropy (or temperature)

than does the maximum in susceptibility. The observed remanent magnetizations are always very small (being only a few per cent of the magnetization at the limits of the cycle, which in turn is only a few per cent of the saturation magnetization) and tend to support the argument against the ordered state being ferromagnetic.

Coincident with the rapid increase of χ towards its maximum and the appearance of hysteresis effects, the a.c. susceptibility is found to become complex (see § 3.2). The out-of-phase component of the susceptibility χ'' increases continuously with decreasing entropy in the cases of cobalt ammonium sulphate (Garrett 1951 a) and ferric ammonium alum (Steenland, de Klerk, Potters and Gorter 1951) but goes through a very sharp maximum just below the Curie point in the case of chromic methylammonium alum (Hudson and McLane 1954). There is conflicting evidence on the behaviour of chromic potassium alum: de Klerk, Steenland and Gorter (1949) report a maximum, while Ambler and Hudson (1954 a) could detect no maximum but only an inflection at roughly the same entropy.

Cooke and Hull (1937) investigated the variation of χ'' with temperature and with frequency ν in ferric ammonium alum (above the Curie point) by observing rates of warming in an a.c. field, using frequencies of 328 and 455 c/s. Making use of the concept of a relaxation time τ within which the spins take up their stable orientation after a change has taken place in H, these authors found that τ varied with temperature according to the relation $\tau=2{\cdot}3\times10^{-7}\exp(-T/0{\cdot}035)$ sec within the range of measurement, 0·04° to 0·12°K. Interpreting the results of MacDougall and Giauque (1936) for gadolinium nitrobenzene sulphonate heptahydrate in a similar way, they find $\tau=5\times10^{-7}$ sec at 0·25°K. This is comparable with the values obtained for ferric ammonium alum. The evidence points to τ representing the time constant for mutual interaction between the spins. The heating effect was found to be proportional to the square of the frequency, implying, $\chi''\propto\nu$ and ruling out any explanation of the phenomenon on the basis of simple magnetic hysteresis. On the other hand, Shire and Barkla (1939), using the same salt, found the heating to be proportional to frequency in the range 75–400 c/s at temperatures in the region of 0·1°K.

De Klerk, Steenland and Gorter (1949) measured χ'' directly in chromic potassium alum for the frequencies 225, 375 and 525 c/s but the measurements were not sufficiently reproducible to permit conclusions to be drawn as to the frequency dependence. By using the frequencies 225 and 525 c/s alternately in the same run, however, it was deduced that, in the range 0·01°–0·05°K, the heating was proportional to $\nu^{1\cdot7}$, in fair agreement with the results of Cooke and Hull for ferric ammonium alum.

The fact that χ'' is closely associated with the appearance of a Curie point strongly suggests that it has its origin in the magnetic interaction forces which produce the transition. In contrast to the spin–spin relaxation observed at higher temperatures (see, for example, Cooke 1950) the relaxation time is temperature dependent. It is not impossible that the two phenomena are basically the same: in the low temperature case we have a situation which is more difficult to examine analytically due to kT becoming comparable with the interaction energy, to the anisotropy of electric interactions, and to the possible development of domain structure.

4.3. *Specific Heat Anomalies*

Any process within a substance which brings about a change of entropy will contribute to the heat capacity. As the temperature is lowered redistribution of the ionic magnetic moments over the allowed energy levels, split by the crystalline field,

will occur giving rise to a Schottky-type anomaly. The removal of the final $R \ln 2$ of electronic entropy occurs through magnetic (or exchange) interaction and this, being a co-operative effect, results in a very rapid decrease in S with a consequent sharp-peak anomaly. A third source of entropy lies in the splitting of nuclear magnetic energy levels by the magnetic field of the electrons surrounding the nucleus. This entropy amounts to $R \ln (2I+1)$ where I is the nuclear spin.

These three effects may or may not overlap, depending upon the relative magnitude of the interactions. For example, in the case of chromic potassium alum the magnetic and crystal field anomalies are centred on 0·01°K and 0·10°K respectively, and are thus clearly separated. The relative abundance of ^{53}Cr, the only stable isotope having non-zero spin, is only 9% and the hyperfine splitting is very small (Bleaney and Bowers 1951) so that the hyperfine structure contribution to the specific heat is negligible. In contrast, in the case of manganous ammonium sulphate, all three interactions are important and their contributions to the specific heat overlap.

When overlap does occur it is frequently possible to determine the various contributions by making measurements upon crystals wherein some of the paramagnetic ions have been replaced by suitable diamagnetic ions. This technique has been used by Benzie and Cooke (1949) in the liquid helium region and by Garrett (1950 a, b) and Malaker (1951) below 1°K. At ' high ' temperatures each interaction contributes a term proportional to $1/T^2$ to the specific heat, and may be determined by extrapolating the results to infinite dilution. The measured specific heat, less the hyperfine structure and Stark contributions, represents the combined effect of the dipole–dipole and exchange interactions. The dipole quantity is usually calculated from the Van Vleck formula (corresponding to equation (11)) and the remainder ascribed to exchange. As discussed in § 2.1 (*d*), the nuclear specific heat may also be calculated from paramagnetic resonance data and the values so obtained are generally in good agreement with those obtained by the dilution method. Care must be exercised in operating on the assumption that Stark splittings are unaffected by dilution. Kurti (private communication) has pointed out that this would be very wrong in the case of ferric rubidium alum, for example.

4.4. *Free Radicals and F-Centres*

The magnitude of the interactions in a paramagnetic determines the temperature of the rapid fall in entropy, and hence the temperature which may be reached by adiabatic demagnetization (see figure 1). Such interactions should be smallest in a spin one-half paramagnetic having a high degree of magnetic dilution. For this reason, some investigations have been undertaken recently into the magnetic behaviour of free radicals (van den Handel 1952, Gerritsen *et al.* 1954) and colour centres in alkali halides (Heer and Rausch 1953).

The evidence to date is not encouraging, however, since it appears that in these situations large exchange interactions (Hutchinson 1949, Holden *et al.* 1950, 1951, Schneider and England 1951) and significant hyperfine interactions (Pake, Townsend and Weissman 1952, Lloyd and Pake 1954) are present. Adiabatic demagnetizations have been made with diphenyl-picrylhydrazyl (Ambler and Hudson 1954, unpublished) but merely to confirm the existence of a susceptibility maximum in the upper tenths-of-degree range. Above 2°K this substance was found to obey a Curie–Weiss law with $\Delta \simeq -0{\cdot}06$ degree; below 2°K, $\chi(T-\Delta)$ decreased rapidly with falling temperature.

4.5. *Antiferromagnetism*

At or close to the Curie point the susceptibility reaches a maximum value. If measurements are made at lower entropies in superimposed steady fields H it is found that the isentropic differential susceptibility $(\partial M/\partial H)_S$ first increases with increasing H and then goes through a maximum. If one plots such results as a lattice of $(M, S)_H$ curves, the sealso show a maximum, provided that H does not exceed a certain critical value, H_c^0; the latter varies considerably from substance to substance. The locus of the maxima constitutes a phase boundary. Within the boundary, $(\partial M/\partial S)_H$ is positive, corresponding to a negative magnetocaloric effect, as may be seen at once from the thermodynamic relation of equation (1). That is to say, isentropic application of a magnetic field results in a cooling. Outside the boundary the reverse situation holds and the substance is paramagnetic, with a positive magnetocaloric effect. The behaviour within the boundary is readily explained for the case of antiferromagnetism:

We may write
$$\left(\frac{\partial T}{\partial H}\right)_S = -\frac{(\partial S/\partial H)_T}{(\partial S/\partial T)_H}$$

and since $\partial S/\partial T$ is always positive, a negative $(\partial T/\partial H)_S$ corresponds to a positive $(\partial S/\partial H)_T$, i.e. isothermal application of a magnetic field *increases* the entropy or disorder. Such an effect would be expected in an antiferromagnetic since the effect of applying a small field will be partially to disrupt the antiparallel alignment of the spins.

The phase-boundary feature appears to be common to all low temperature paramagnetics. It has been reported by Kurti (1951) for a compressed powder 3–1 ellipsoid of ferric ammonium alum, by Garrett (1951 a) for single crystal sphere of cobalt ammonium sulphate along the K_1-axis only, by Ambler and Hudson (1954 b) for a single crystal sphere of chromic methylammonium alum measuring along a cubic axis, and may also be derived from the results of Steenland (1952) for a single crystal sphere of chromic potassium alum.

Garrett first drew attention to the phase boundary and discussed his results on the cobalt Tutton salt in the light of a theory of antiferromagnetism, due to Sauer and Temperley (1940), based on dipole–dipole interaction. This predicted a transition curve quite similar to that observed experimentally. Comparison was made on the basis of a $(T, H)_S$ diagram, which is essentially the same as a $(M, S)_H$ diagram by virtue of equation (1). The transition curve is roughly parabolic in shape and along it the field (termed ' critical field ' and denoted by H_c) varies between zero at the normal Curie point T_c and a limiting value H_c^0 at the absolute zero.

The Sauer–Temperley theory, designed for the alums (face-centred cubic lattice) and with the simplifying assumptions of no crystalline field effects and a unique direction of alignment, requires $T_c = 3{\cdot}6\ N\mu^2/k$ and $H_c^0 = 1{\cdot}8\ N\mu$ where N is the number of ions per cm^3 and μ is the magnetic moment of each ion. The crystalline electric field in the cobalt Tutton salt has tetragonal symmetry and there are two sets of magnetic ions with different tetragonal axes. Garrett took for his value of the critical field H_c^0 the component along a tetragonal axis of the ' external critical field '. The theoretical and observed values of T_c and H_c^0 were respectively 0·065°K and 0·084°K, 160 oersteds and 340 oersteds.

For chromic methylammonium alum, Ambler and Hudson (1954 b) obtained $H_c^0 = 120$ oersteds, which has a component of 70 oersteds along each of the trigonal

axes—the presumed directions of alignment. The theoretical value is 100 oersteds. The value of T_c for this salt is 0·02°K (Gardner and Kurti 1954); the theoretical value is 0·04°K. In contrast, T_c for chromic potassium alum is either 0·004°K or 0·011°K (see § 3.7) and H_c^0 appears to be of the order of 15 oersteds.

Dipole–dipole interaction is present and is of roughly the same order of magnitude in all low-temperature paramagnetics, but the above results make it extremely unlikely that it is solely responsible for the co-operative transitions. If this were so, the Curie points of these two chromic alums should be very nearly the same, since the magnetic ion is the same and the inter-ionic distance differs by only 3%. Garrett (1951 b) has been able to show that an extension of Van Vleck's theory of direct exchange interaction (applicable to a simple cubic or body-centred cubic arrangement) leads to results remarkably similar, in main features, to those obtained by Sauer and Temperley for the case of dipole–dipole coupling. Direct exchange, however, can almost certainly be excluded, theoretically by reason of the high dilution, and experimentally on the basis of no detectable Curie–Weiss Δ, but indirect exchange can be admitted. The existence of indirect exchange has been invoked to account for otherwise unidentifiable contributions to the heat capacity, but the theory is not sufficiently well developed to be applied to co-operative phenomena. (See, however, the extension by Gorter and Haantjes (1952) of Néel's phenomenological theory.)

4.6. *Thermal Conductivity*

The thermal conductivity k of paramagnetic salts is an important factor in research at very low temperatures. It is found to be very small and thus any stray heat-leak to a cooled specimen will tend to raise the temperature of the surface preferentially and give rise to an inhomogeneous temperature distribution. It is difficult to apply a satisfactory correction for this effect, and the ' useful time ' after a demagnetization is severely limited, being inversely proportional to the heat-leak and probably of the order of a few minutes for a typical case of 2–3 erg min^{-1} cm^{-2}. Again, k partly determines the degree of thermal equilibrium between a paramagnetic salt used as a coolant and a second substance which is being cooled through thermal contact (see § 5.2).

Relatively few measurements have been made of k for paramagnetic salts; the data available have been summarized and discussed by Berman (1953). The standard experimental methods are of little use below 1°K owing to the difficulties of establishing thermal contact between the salt specimen and heaters or thermometers, and of minimizing heat-leaks. The usual procedure has been to measure a mean conductivity by observing the rate of temperature equalization of the ends of paramagnetic crystals demagnetized from inhomogeneous fields. An average magnetic susceptibility is measured for each end of the crystal, from which one derives an average T^* and hence an average T. The rate of heat flow is calculated from heat capacity data, a correction being applied for the heat-leak. Considering the inaccuracies involved at each step of this treatment, it may be appreciated that no more than an order-of-magnitude determination can be expected.

Measurements were made by Kurti, Rollin and Simon (1936) on cylinders, 7 mm in diameter, cut from crystals of ferric ammonium alum (23 mm long) and chromic potassium alum (30 mm long). Values for k (in cal deg^{-1} cm^{-1} sec^{-1}) were obtained as follows: for the ferric alum, $k = 0{\cdot}6 \times 10^{-5}$ at $T^* = 0{\cdot}07$ degree and $1{\cdot}8 \times 10^{-5}$ at 0·10 degree; for the chromic alum $k = 2{\cdot}3 \times 10^{-5}$ at $T^* = 0{\cdot}18$ degree.

Bijl (1949) measured k for chromic potassium alum between 1·4° and 3·9°K by a conventional heating method, determining the temperature gradient by the susceptibility method; k was found to be proportional to a power of the temperature of about 2·3. It was also found that the conductivity depended on the rate at which the specimen was cooled below 70°K. Garrett (1950 c) examined the same salt between 0·16° and 0·29°K by a method not very different from that of Kurti *et al.* In this region k varied as T^3 within the experimental accuracy and seemed to fit on to an extrapolation of the ' slow cooling ' curve of Bijl.

The thermal conductivity may be expressed in terms of a mean free path l (Debye 1914), the distance which a lattice wave travels before its intensity is reduced to 1/e of its initial value. Retaining the concept of a mean free path when discussing heat transfer in terms of phonons, the Debye expression $k=\frac{1}{4}Cvl$ is modified to $k=\frac{1}{3}Cvl$ where C is the lattice specific heat per unit volume and v is the velocity of sound. The mean free path should increase with decreasing temperature (Peierls 1929) and at very low temperatures, l should be determined by the size of the specimen being investigated, or by that of the grains of which it may be composed, and k should vary as C, i.e. as T^3. Although the latter type of variation was observed by Garrett, his results indicate that l was about 0·5 mm while the crystal diameter was 15 mm. Berman points out that this suggests there are boundaries within the crystal sufficiently definite to scatter phonons and limit the mean free path to 0·5 mm. The results of Kurti *et al.* on ferric ammonium alum indicate $l\sim0{\cdot}25$ mm, whereas the crystal diameter was 7 mm.

The earliest measurements showed that specimens in the form of compresses of finely ground crystals exhibited a thermal conductivity lower by a factor ten than that of a single crystal. Van Dijk and Keesom (1940) obtained the relation $k=0{\cdot}00312\times(T^*)^3$ calorie units for the conductivity of compressed, powdered ferric ammonium alum in the range $T^*=0{\cdot}04$–$0{\cdot}3$ degree. At 0·1°K this gives a value about one sixth that of the single crystal value of Kurti *et al.* The density of the specimen was very close to that of the single crystal. Hudson (1949 b) investigated a powder specimen of the same substance in the range 0·10–0·22°K. At 0·10°K the conductivity was about fifty times less than the single crystal value; the filling factor was again very close to unity and the size of the crystallites was estimated to vary between 10^{-3} and 10^{-2} cm. Berman points out that these results suggest a mean free path of the order of the size of the crystallites.

In cooling a non-paramagnetic substance indirectly by means of the magnetic cooling method, a frequently used technique is to establish thermal contact between the substance and paramagnetic refrigerant by immersing both in liquid helium. It may be seen, therefore, that as far as the properties of the salt are concerned, heat transfer should be made quite efficient by using the salt in finely powdered form, the transmission area increasing as d^2 while k decreases as d, where d is the average linear dimension of the grains. The temperature discontinuity at the liquid-salt interface is another important consideration; this is taken up in § 5.2.

§ 5. Thermal Insulation and Thermal Contact below 1°K

5.1. *Thermal Insulation*

Heat influxes to the salt arise through absorption of radiation, condensation of helium gas, conduction along supports, and dissipation of energy stored in vibrations of the specimen and supports. By taking suitable precautions the total heat influx

can be brought down to 20–50 erg min^{-1}, which is small enough to allow most salts to remain below 1°K for many hours.

Radiation falling on the salt should not be allowed to come from surfaces at temperatures above a few degrees absolute.† The walls of the salt chamber are accordingly made opaque and radiation traps are included in pumping lines connected to the chamber.

Although after demagnetization the pressure of helium in the salt chamber must become very small on account of the 'gettering' action of the salt, there remains a layer of helium absorbed on the walls, part of which gradually desorbs and condenses on the salt. This may be removed before demagnetization by prolonged pumping, but when metal cryostats are used and the starting temperature is very low (e.g. 0·9°K) this may take an inconveniently long time. The effect may be reduced considerably by a 'degassing' technique when precise knowledge of the initial temperature, i.e. T_i, is unimportant. The cryostat is pumped at a temperature of about 1·5°K for some time, then just before demagnetizing the temperature is reduced as low as possible. In any case, when the temperature of the salt rises above about 0·5°K, helium begins to desorb from the salt, the vacuum in the salt chamber consequently deteriorates and the heat-leak increases considerably. This makes it difficult to carry out experiments between 0·5°K and 1°K.

Some of the difficulties inherent in the use of exchange gas can be overcome by using a metallic link between the salt and the helium bath through which the heat of magnetization can flow. Heat contact must be broken before demagnetization, however, and this may be done either mechanically (Collins and Zimmerman 1953) or by a superconducting switch (Heer and Daunt 1949). The ratio of the thermal conductivity of a pure superconductor in its normal and superconducting states k_n/k_s increases strongly for decreasing temperature and for tin or tantalum, for example, is about 50 at 0·5°K (see § 6.1 (*c*)). The switch is operated by means of a magnetic field greater than the threshold value. Additional information can be found in papers by Gorter (1948 a), Mendelssohn and Olsen (1950 a), Steele and Hein (1952), Nicol and Daunt (1952), Croft, Faulkner, Hatton and Seymour (1953), and de Vries and Daunt (1954).

The support for a salt specimen should be designed with two points in mind: the heat conducted along it must be small and the system should not be free to execute large vibrations. By using fine enough supports of low thermal conductivity material there is usually not much difficulty in reducing the heat conduction to a tolerable value. For example, if one extrapolates data by Berman (1951) for the thermal conductivity of glass, a spill 1 mm diameter, of 0·1 mm wall thickness and 5 cm long would conduct roughly 5 erg min^{-1} from a helium bath at 1°K to the salt. High warming up rates due to vibration in the system (see, for example, Hull, Wilkinson and Wilks 1951, Darby, Hatton, Rollin, Seymour and Silsbee 1951) have been attributed to frictional losses, either internally in the suspension material or superficially at the points of contact with the salt, and steps should be taken to insulate the apparatus from likely sources of vibration such as vacuum pumps. A more rigid suspension for the salt often helps, but this may result in an increased heat-leak due to conduction. The fact that good thermal contacts at very low

† At 300°K, 90°K and 20°K, $2{\cdot}8\times10^{7}$, $2{\cdot}2\times10^{5}$ and $5{\cdot}5\times10^{2}$ erg min^{-1} respectively are emitted per cm^2 of a black body, and the spectrum of the radiation has a maximum intensity for wavelengths λ about 10μ, 30μ and 150μ respectively. In the infra-red region, radiation for which $\lambda > 5\mu$ is not transmitted by ordinary glasses.

temperatures are extremely difficult to make may be exploited and a high thermal resistance introduced at the point of contact between salt and support. There is also the possibility of interposing a guard ring at some point along the support, consisting of a piece of paramagnetic salt, having a large specific heat at a temperature of about 0·2°K, which is demagnetized at the same time as the specimen.

Although a heat influx of 20–50 erg min^{-1} can be reproducibly achieved, this is not the case for smaller heat influxes, and there are very few such cases reported (Cooke and Hull 1942, Darby *et al.* 1951, Malaker 1951). In fact one of the main technical problems still to be solved is that of designing a system into which there is, reproducibly, a very small heat influx, say 1 erg min^{-1}. This is very important if appreciable progress is to be made in the field of absolute temperature determinations at the lowest temperatures, and if nuclear demagnetizations are to be carried out successfully (see § 7.2).

5.2. *Thermal Contact*

Below about 0·7°K thermal contact can only be made in two ways, (*a*) mechanically, by pressing the substance to be cooled to the salt, (*b*) through the intermediary of liquid (or solid) helium.

(*a*) Cooling by mechanical contact was first used by Kurti and Simon (1934) to discover new superconductors. The superconductor was finely divided, mixed with powdered ferric ammonium alum, and the mixture compressed to form a very compact pill. The method has been modified by many workers to study a variety of problems, and individual cases are discussed in § 6. Mendoza (1948) mixed a lacquer with the powdered salt to improve the contact between the salt and copper fins embedded in the mixture. He measured the equilibrium times for heat transfer between two such pills of chromic potassium alum, the copper fins being connected by a copper rod some 10 cm long. He found that at temperatures down to about 0·15°K the main thermal resistance was that of the contact between the salt and the fins, the rate of heat flow through which obeyed the relation,

$$dQ/dt=(\beta A/3)(T_1^3-T_2^3) \qquad \ldots\ldots(16)$$

where A was the area of contact, T_1-T_2 the temperature difference across the contact and β a constant $=3\times10^{-5}$ watt cm^{-2} deg^{-3}. This relation was subsequently verified by Goodman (1953) who obtained, however, a somewhat larger value for β.

(*b*) Two methods have been developed using liquid helium:

(*α*) The sealed high pressure capsule method was devised by Kurti, Rollin and Simon (1936) and has been used by Kurti and Simon (1938 a), Shire and Allen (1938), and Hull, Wilkinson and Wilks (1951). A similar but less efficient method has been used by de Haas, Casimir and van den Berg (1938). The substance under investigation is placed in a thick-walled metal capsule together with powdered paramagnetic salt. The capsule is then filled at room temperature with helium gas at a pressure of 120 atmospheres and sealed off. The procedure has been described in detail by Hull (1947). The volumes of salt and container are so chosen that when the capsule is cooled and the helium liquefies, the salt and specimen are completely covered with liquid.

(*β*) The open capsule method was devised by Hudson, Hunt and Kurti (1949) and has been used in many different experiments which will also be described in detail in § 6. The apparatus is shown in figure 6. The open container A holding the liquid helium, salt and specimen is suspended by the capillary tube C

inside a vacuum case suspended in the helium cryostat. In the original experiment C was 7 cm long with internal diameter 0·2 mm and external diameter 0·5 mm. The main source of the heat influx is due to the helium, which, after creeping up the walls of the connecting tube C, evaporates in the warmer parts of the apparatus and condenses back into the container. The action of the capillary is twofold; it limits the amount of film creeping out by providing only a small transfer perimeter, and it limits the amount of gas condensing back by providing a big flow resistance. By this technique the heat influx to A was kept down to 300 erg min^{-1} and by using a double-container system, in which the upper one acted as a guard-ring for the lower, the leak was reduced to 50 erg min^{-1}. An attempt was made to keep the pressure at the top of C as low as possible by connecting it to a wider tube and pumping on the latter with a fast diffusion pump. It has been found, however, that the pumping of C does not significantly reduce the heat influx into A (Hudson 1949 b), but that the heat influx increases considerably with increasing cryostat temperature. It would seem therefore that the pressure at the top of C must always be that corresponding to the vapour pressure of helium at the cryostat temperature and that the essential points of good design are to have as low a cryostat temperature as possible and a big flow resistance in C. (A similar method has been used by

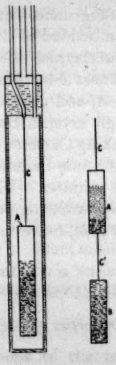


Figure 6. Open-container arrangements for cooling liquid helium below 1 °K (*after* Hudson, Hunt and Kurti 1949).

de Klerk (1946), who inserted a valve at the bottom of the tube connected to the capsule.)

It becomes more and more difficult to use liquid helium as a heat transmitting agent at very low temperatures on account of the falling off in the thermal conductivity. (The measurements of Kurti and Simon (1938 a) between 0·2°K and 0·5°K give $k=2{\cdot}5\times10^{-2}T^3$ cal cm^{-1} sec^{-1} deg^{-1}.) Furthermore, with liquid helium there is another thermal resistance to be considered, namely the contact resistance between the liquid and a solid surface, discovered by Kapitza (1941). From measurements made above 1°K (Kapitza 1941, Osborne 1951, Beenakker, Taconis, Lynton, Dokoupil and van Soest 1952) one may take heat flow through such a contact across which there is a temperature difference ΔT as being

approximately given by $dQ/dt = T^3 \Delta T/32$ cal cm^{-2} sec^{-1}. Measurements of Ambler (1953) indicate that this temperature relationship can be extrapolated at least down to 0·1°K. This means that contact resistance is extremely important when liquid helium is used as the transmitting agent; indeed it has the same effect as a layer of liquid helium 8 mm thick. Theories of this contact resistance have been developed by Kronig and Thellung (1950), Gorter, Taconis and Beenakker (1951), Khalatnikov (1952).

The use of solid helium for establishing thermal contact within an open capsule has been suggested by Swenson (see Wilkinson and Wilks 1951). Webb, Wilkinson and Wilks (1952) found that at temperatures below about 0·2°K, its normal conductivity at densities obtainable with quite moderate pressures (~50 atmospheres) is about ten times that of the liquid and furthermore, on the basis of all the above theories, one would expect a smaller surface thermal resistance effect than is encountered with the liquid.

Another important factor in the design of apparatus for cooling by contact below 1°K is the thermal diffusivity of paramagnetic salts. The thermal conductivity varies as T^3 (see § 4.6) but the specific heat varies with T in a complicated way and a rigorous analysis of heat flow is impossible. At any given temperature, however, one can make a rough computation.

For example, consider an arrangement where one is establishing thermal contact by means of copper fins embedded in the salt, the separation of adjacent fins being l. One may estimate the thermal time constant τ for a slab of salt (say chromic potassium alum) of thickness l by supposing that there exists a small temperature inhomogeneity across it, and using the relation $\tau \simeq l^2/\delta\pi^2$, where δ is the thermal diffusivity. Using single crystals, the smallest one could conveniently make l would be about 0·5 mm. Taking Garrett's value for the thermal conductivity $k = 1{\cdot}25 \times 10^{-2} T^3$ cal cm^{-1} sec^{-1} deg^{-1}, one finds that at 0·1°K, $\tau \simeq 0{\cdot}1$ sec, and at 0·01°K this is increased to several minutes. The situation would be somewhat worse if one used salt in the form of a powder compress since the thermal conductivity is about ten times smaller; in this case, however, one could conveniently make l still smaller.

§ 6. Experiments below 1°K

6.1. *Properties of Metals*†

(*a*) *Electrical conductivity.*

As the temperature of a metal is lowered the electrical resistivity is found to fall until in the liquid hydrogen or liquid helium region a levelling off occurs. This residual resistance is probably due to the mean free path of the electrons reaching a limit determined by the presence of various types of impurities, physical and chemical. At still lower temperatures the resistivity is sometimes found to drop abruptly to zero (superconductivity); in other cases the resistivity has occasionally been found to rise again, thus demonstrating the phenomenon of a resistance minimum.

The latter has not been investigated very much below 1°K due mainly to the inherent experimental difficulties. De Haas, Casimir and van den Berg (1938) studied the behaviour of a gold wire down to 0·4°K, their limit of good thermal

† *Note added in proof.* Some measurements have been reported recently by J. Rayne (*Phys. Rev.*, 1954, **95**, 1428) on the specific heats below 1°K of a number of metals.

equilibrium. The wire was mounted in a glass tube (containing helium gas for heat transfer) enclosed in a quantity of ferric ammonium alum. A phosphor bronze wire, also in the tube, was used as a resistance thermometer. The resistivity of the gold was found to increase continuously with falling temperature and the measurements suggested that the rate of rise would become extremely rapid below 0·2°K.

Mendoza and Thomas (1951) investigated gold, silver and copper down to about 0·1°K. The paramagnetic salt was pressed around copper fins connected by a copper wire to a copper former which supported the wire specimen under investigation. Insulation was provided by a thin coating of bakelite varnish on the former, and electrical contacts to the wire were made through thin superconducting (high thermal resistivity) leads and the copper fins. All three metals showed a rise in electrical resistivity, becoming very rapid at the lowest temperatures. The minimum resistance decreased when the physical purity of a specimen was enhanced by annealing. Mendoza and Thomas found that the low-temperature rise in resistance was steeper the greater was the value of the minimum resistance.

Further measurements were made by Thomas and Mendoza (1952) on magnesium and molybdenum. The resistivity of the latter reached a constant value at the lowest temperatures, after rising exponentially with decreasing temperature over a considerable range. The electrical resistivity of gold was measured down to 0·006°K by Croft *et al.* (1953) and found to obey a relation of the form $R=a \log (b/T)$. Two methods were employed in these determinations, the open-capsule technique and a two-stage demagnetization technique (see § 7.1).

The phenomenon of the resistance minimum is not well understood and it is possible that such minima may be produced by different mechanisms in different metals. Various explanations have been put forward, e.g. scattering of electrons by unfilled d-shells of impurity atoms (Gerritsen and Korringa 1951), limitation of the electron mean free path due to the physical size of the specimen, and especially (MacDonald 1951) a size effect due to internal boundaries, a rearrangement of the lattice to cause a gap at the top of the electron distribution and resultant semiconductor behaviour (Slater 1951). MacDonald (1952) has recently reported the results of a series of measurements on a variety of alloys—not below 1°K—and points out that no existing theory is adequate to account for these results.

(*b*) *Superconductivity.*

One of the earliest applications of the magnetic cooling method was the discovery of new superconductors with transition temperatures below 1°K (Kurti and Simon 1934, 1935 a). By embedding the metal under investigation in a powdered paramagnetic salt and observing the change of susceptibility during warm-up, a transition may be detected by the appearance of a large discontinuity in χ due to the disappearance of the superconductor's diamagnetism. Transitions were observed for cadmium, hafnium and zirconium.

For investigating the detailed behaviour of the superconductor it is best to employ the technique of Goodman and Mendoza (1951) in which the metal is separated from the salt by means of a copper rod. These workers measured the temperature dependence of the threshold magnetic field in aluminium, cadmium, gallium and zinc, and confirmed the parabolic relation for these substances. Superconducting transitions in osmium and ruthenium have been observed by Goodman (1951).

The method of mixtures has been employed by Daunt and co-workers to obtain critical field data for titanium (Daunt and Heer 1949 a), aluminium and zinc (Daunt and Heer 1949 b), and zirconium, hafnium, cadmium and titanium (Smith and Daunt 1952). In the latter work attention was paid to the considerable effects of impurities and strain, which were further emphasized by subsequent measurements on a well-annealed, pure specimen of titanium by Smith, Gager and Daunt (1953). These gave values of T_c and $(dH_c/dT)_{T_c}$ considerably smaller than those obtained with the specimen investigated earlier. Discrepancies between the results of different workers for the same metals are very probably due to varying degrees of strain in the specimens, especially in hard materials such as Hf, Zr and Ti. The possibility of errors in temperature measurements, however, cannot be entirely discounted.

Calorimetric measurements are difficult to make by reason of the extremely small specific heats of metals in this region. Samoilov (1952) has investigated cadmium between 0·3 and 0·9°K; above the transition point he found $C = 1{\cdot}70 \times 10^{-4} T + 464(T/300)^3$ cal mol^{-1} deg^{-1}, the linear term being in fair agreement with the figure derived from magnetic measurements (Samoilov, 1951), and below T_c the specific heat varied as T^3 (as in most superconductors).

(*c*) *Thermal conductivity.*

The thermal conductivity of metals, both normal and superconducting, at very low temperatures is not only of great fundamental interest but it is also of considerable practical importance in applications of the magnetic cooling method. As discussed more fully in § 5.2 (*a*), a convenient and satisfactory way of cooling a second substance is by means of a ' thermal link ' of, for instance, copper to the paramagnetic salt. The marked decrease in k which occurs when a pure metal enters the superconducting state forms the basis of the ' superconducting heat switch ', which is discussed in §§ 5.1 and 7.1.

The first measurements below 1°K were made by Heer and Daunt (1949) who investigated tin and tantalum down to 0·2°K by using a metal specimen as a link between two ' thermometers ' of chromic potassium alum. For both metals the thermal conductivity in the superconducting state k_s was found to be much smaller than that in the normal state k_n, the ratio k_n/k_s in the region of 0·5°K being about 60 for tin and 40 for tantalum. For tin, $k_s = 1{\cdot}1 \times 10^{-2} T^3$ watt unit. By extrapolation of these results it was estimated that for tin at 0·01°K the ratio k_n/k_s would be of the order of 10^5.

The heat conductivity of lead down to 0·4°K has been measured by Olsen and Renton (1952). Measurements were made for the superconducting and ' intermediate ' states, but the normal-state conductivity was too large to permit measurements on the same specimen. Heat was supplied electrically at one end of the specimen, the other end being embedded in a ' pill ' of compressed paramagnetic salt which served as a heat sink. The temperature gradient was measured by means of carbon resistance thermometers (§ 3.5), attached at two intermediate points, a method which avoids uncertainties due to contact-resistance encountered (especially with magnetic thermometers) when the temperatures at the ends of the specimen are measured. Up to 0·9°K, k appeared to be proportional to T^3, with a stronger dependence on T above this point. In the intermediate state the rise in thermal resistivity and marked hysteresis effects observed by Mendelssohn and Olsen (1950 b) occurred. Similar results were obtained by Mendelssohn and Renton

(1953) for tin, indium, thallium, columbium, tantalum and aluminium. The conclusion was reached that in the T^3 region, only the crystal lattice is contributing to the conductivity, although the numerical values of k were five to ten times smaller than those predicted theoretically for ideal single crystals.

Goodman (1953) published data obtained for five tin specimens of widely differing purity. Here again the specimen rod joined two pills of paramagnetic salt, via copper inserts as described by Mendoza (1948). The results were discussed in terms of the separate contributions to the conductivity from the electrons and from the lattice. The former was found to be in agreement with the theoretical predictions of Heisenberg (1948).

The thermal conductivity of copper down to 0·25°K was studied by Nicol and Tseng (1953) using the same technique as Heer and Daunt and applying a correction for contact resistance. A linear relation, $k=\alpha T$, was verified with $\alpha=1{\cdot}76\text{w cm}^{-1}\text{ deg}^{-2}$, which held over the temperature range 0·25–4·2°K. Previous results in the region 1·2–4°K give for α the values 0·6 (Allen and Mendoza 1947), 0·58 (Andrews, Spohr and Webber 1950), 4·7 (Berman and MacDonald 1952) and 2·9 (Mendelssohn and Rosenberg 1952).

6.2. *Properties of Helium*

(*a*) *Liquid helium.*

The striking properties of liquid helium which are manifested below 2·2°K (i.e. by liquid helium II) such as superfluidity, enhanced heat transport, and second sound, are strongly temperature dependent. How they are modified by a progressive lowering of the temperature below 1°K is a subject which has lately become of great interest to both experimental and theoretical physicists.

The experiments of Pellam and Scott (1949) conclusively demonstrated a marked rise in the second sound velocity u_2 below 1°K and resolved the conflict between the theoretical predictions of Tisza (1940) and Landau (1941) in favour of the latter. Higher velocities at still lower temperatures were observed by Atkins and Osborne (1950) and appeared to confirm the prediction of Landau that u_2 would tend towards a limiting value of $u_1/\sqrt{3}$ where u_1 is the velocity of sound. More recently de Klerk, Hudson and Pellam (1953, 1954) measured the variation of u_2 with temperature over the range 0·02° to 1°K. Their results, illustrated in figure 7, demonstrated the occurrence of a sharp rise in u_2 at 0·7°K followed by a tendency to flatten off at the ' Landau velocity ' at 0·5°K and a subsequent rise as the temperature was lowered further. The indications are that the value of u_2 approaches that of u_1 at very low temperatures. The shape of the received pulses becomes very much spread out below 0·6°K and gives rise to the speculation (Pellam and Scott 1949) that at the lowest temperatures one may have to do with a return to normal diffusive heat flow. In the Landau terminology, the contribution of roton (high temperature) excitations to ' the normal fluid ' density varies exponentially with temperature and, below a few tenths of a degree Kelvin, phonons will be entirely responsible for the thermal transport properties. Explanations of the observed pulse broadening and anomalous increase in u_2 towards the value of u_1 at very low temperatures have been given in terms of phonon mean-free-path effects (Atkins 1953, Mayper and Herlin 1953, de Klerk, Hudson and Pellam 1953) and Ziman (1954) has been able to account for the observed shape of the received pulses in a quantitative analysis.

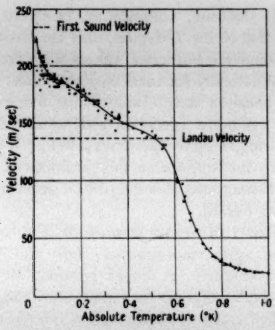


Figure 7. Velocity of second sound in liquid helium II as a function of temperature (*after* de Klerk, Hudson and Pellam 1954).

The thermal conductivity of liquid helium below 1°K has only been measured to date by rather rough methods. Kurti and Simon (1938 a) used two sealed capsules each containing about 1 gramme of ferric ammonium alum and helium gas under high pressure, and connected by a capillary, 18 mm long with 0·5 mm internal diameter. In the range 0·2–0·5°K, k appeared to vary as T^3 and the authors concluded that at such temperatures the conductivity was entirely 'lattice-type' and no longer enhanced by the superflow properties. During his measurements on the specific heat of copper potassium sulphate, de Klerk (1946) obtained some values for k in the range 0·2 to 0·9°K. Two glass spheres 36 mm in diameter were connected by a glass tube 10·5 cm long and 3·2 mm internal diameter and for this part of the experiment were both filled with chromic potassium alum. The apparatus was provided with a valve to reduce heat influx due to film flow. The rate of temperature rise of the lower sphere after demagnetization provided a measure of the heat transfer through the liquid filling the connecting tube. The measurements showed a rather abrupt increase in the rate of variation of k with T at about 0·55°K, k apparently being roughly proportional to T^5 above this temperature. The interpretation was given, however, in terms of a 'mean heat conductivity' defined by heat flux divided by average temperature gradient. Since k is obviously a strong function of T, however, one could better write $k=aT^n$ and this leads to (approximately) a T^3 dependence at the higher temperatures. In a similar way, one may deduce from the measurements at the lower temperatures that the index n is 3, at least in the range 0·25–0·50°K, and the value of a is here about 9×10^5 erg sec^{-1} cm^{-1} deg^{-4}. This is in quite good agreement with the results of Kurti and Simon. In a preliminary account of recent experiments, Fairbank and Wilks (1954) report values of k some ten times greater in the low-temperature region than those above. Below 0·6°K the heat flow was found to be linear with temperature gradient and the index n was found to be less than 3.

The velocity of the liquid helium film has been measured by Ambler and Kurti (1952 b) below 1°K. A double-walled beaker was filled with compressed manganous ammonium sulphate and liquid helium was initially introduced into the inner vessel. To determine the rate at which the liquid flowed out of the inner space into the

annular volume through the agency of the mobile film, visual observations were made on the meniscus in the inner beaker. This was viewed through vertical slits, cut in the salt cylinder, and illuminated by a collimated monochromatic light beam (mercury green). Especial care has be taken to prevent light falling directly on the salt, as this would have caused a prohibitively fast warming-up rate. The results are shown in figure 8; a striking feature is the resemblance which this curve bears to the second sound velocity curve shown in figure 7.

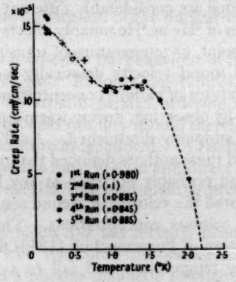


Figure 8. Flow-rate of the liquid helium film as a function of temperature (*after* Ambler and Kurti 1952)

The specific heat of liquid helium at low temperatures was first measured by Pickard and Simon (1939), who found a T^3 dependence below 0·8°K. Keesom and Westmijze (1941) reported a T^6 dependence down to 0·6°K, summarized by $C=0{\cdot}023\ T^6$ cal g^{-1} deg^{-1}. Hull, Wilkinson and Wilks (1951) obtained $C=0{\cdot}024\ T^{6{\cdot}2}$ for the region above 0·6°K, and good agreement with the Landau theoretical curve. In these experiments, measurements below 0·6°K were not possible because of the dominantly large heat capacity of the ferric ammonium alum used for cooling. Kramers, Wasscher and Gorter (1952) obtained more favourable conditions by using copper potassium sulphate and were able to continue their measurements down to 0·25°K. They made use of the open-container technique whereas the earlier work had all been done with a sealed-capsule assembly. Below 0·6°K the specific heat was found to be given by $C=5{\cdot}62\times10^{-3}\ T^3$ cal g^{-1} deg^{-1}, in good agreement with the theoretical value calculated for a phonon gas. Above 0·6°K the specific heat will contain, on the Landau picture, a roton term in addition to the phonon contribution. The theoretical expression for the roton contribution has the form (Landau 1947) $C_r=Af(T)\exp(-\Delta/kT)$ and Kramers *et al.* noted reasonably good correspondence between theory and experiment for $\Delta/k=8$°K. It appears that a choice of 9·6°K is even better (de Klerk, Hudson and Pellam 1954) and this agrees with the second sound velocity results and Landau's original evaluation.

(*b*) *Solid helium.*

The thermal conductivity of solid helium at a density of 0·194 g cm^{-3} was measured by Webb, Wilkinson and Wilks (1952) between 0·3° and 0·7°K. k was found to vary approximately as $T^{2{\cdot}3}$ with a value of about 0·1 w cm^{-1} deg^{-1} at 0·6°K.

Thus in this region, the behaviour of helium is normal in that the thermal conductivity of the solid is appreciably greater than that of the liquid. These results and others obtained by the same authors in the liquid helium region point to a maximum in k occurring at about 0·9°K. The specific heat as a function of temperature down to 0·55°K was also measured for three different densities.

(*c*) *Helium* 3.

In the last year or two, experiments have been reported on liquid ^{3}He. The properties of the light isotope are considerably different from those of ^{4}He and the presence of small quantities of ^{3}He in ^{4}He notably affects the properties of the latter. For example the lambda-point, or temperature of transition from He II to He I, is depressed and the second sound velocity drastically modified. Daunt and Heer (1950) cooled ^{3}He–^{4}He mixtures of various concentrations in an open container and detected the lambda-point by observing, during warm-up, the temperature at which the heat influx to the salt suddenly diminished. An 89% ^{3}He mixture showed a lambda-point of 0·38°K and these authors deduced that pure ^{3}He will not undergo a transition above 0·25°K and probably not even at any lower temperature. King and Fairbank (1953) measured the velocity of second sound between 0·2° and 1·1°K for ^{3}He–^{4}He mixtures of various concentrations. The results were in general agreement with the prediction of Pomeranchuk (1949) that u_2 would go through a maximum with decreasing temperature and fall to a small finite value at 0°K. No significant pulse distortion was observed down to the lowest temperatures, indicating the absence of long-mean-free-path effects such as are encountered in pure ^{4}He.

Roberts and Sydoriak (1954) measured the vapour pressure of pure ^{3}He down to 0·47°K, and the ratio of specific heat at the saturation vapour pressure to latent heat, C_{sat}/L, down to 0·54°K. The vapour pressures agreed within 2% with the equation of Tchen and London (1953). The derived entropy was found to be approaching zero at 0°K, indicating no liquid phase transition at low temperatures. These authors calculated that S/R reaches ln 2 at 0·5°K and falls to about 1·2 at 0·4°K. Measurements of the specific heat have also been reported by de Vries and Daunt (1954) down to 0·57°K and by Osborne, Abraham and Weinstock (1954) down to 0·42°K; in both cases use was made of a carbon resistance thermometer for temperature measurement. The results of Osborne *et al.* were summarized by the expression $C_{sat}=0{\cdot}53+0{\cdot}48\ T$ cal mol^{-1} deg^{-1} which was used, in conjunction with their own vapour pressure data obtained above 1°K, to derive an expression for the entropy. This indicated a value for S/R of 1·35 at 0·42°K. (The decrease of the entropy below the value R ln 2 is of interest in connection with the alignment of the nuclear spins, discussed below.) The three sets of specific heat data agree quite well and show only a small increase of C in the range 0·4–1·0°K.

Weinstock, Abraham and Osborne (1952) measured the melting pressure of ^{3}He down to 0·16°K, extending their earlier investigations in the region above 1°K. Between 0·5° and 1·5°K the data lie on the curve $P=26{\cdot}8+13{\cdot}1\ T^2$ atmospheres. The curve showed a rather sudden flattening just below 0·5°K, the experimentally determined pressure rapidly approaching a constant value of 29·3 atmospheres. The results were later discussed more fully by the same authors (Weinstock *et al.* 1953) and the conclusion reached that no appreciable nuclear alignment occurs in the liquid down to 0·5°K. The Clapeyron–Clausius relation gives $dp/dT=(S_{liq}-S_{sol})/(V_{liq}-V_{sol})$ and since the entropy of the solid probably remains

at R ln 2 down to $T \sim 10^{-7}$°K (Pomeranchuk 1950), any nuclear alignment in the liquid at much higher temperatures than this should be accompanied by a negative sign for dp/dT. Roberts and Sydoriak (1954) pointed out, however, that the existence of a minimum in the melting pressure curve would be obscured by the blocked capillary technique, which was used by Weinstock *et al.* Nuclear alignment should also be evidenced by the entropy of the liquid falling below the value R ln 2 but the computations of Roberts and Sydoriak were in conflict with those of Osborne *et al.* (see above) and the question of alignment at 0·5°K remained somewhat in doubt until the publication of the nuclear resonance results discussed below.

A point of great interest is whether the behaviour of liquid ^{3}He approximates to that of an ideal Fermi–Dirac gas. At low temperatures, the spins of the particles in such a gas should take up an anti-parallel ordering as the entropy is reduced toward zero and the magnetic susceptibility should therefore deviate from the Curie law, finally becoming temperature independent. A study of the nuclear magnetism of liquid ^{3}He therefore offers the most precise and direct information, the measurements of vapour pressure, specific heat and melting pressure providing additional information on this question.

A definitive experiment below 1°K has recently been carried out by Fairbank, Ard and Walters (1954) who measured the susceptibility (down to 0·2°K) by observing the amplitude of a nuclear magnetic resonance signal. The results are shown in figure 9. Here curve A represents the Curie law and curve C the

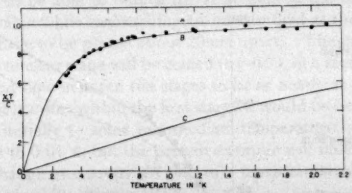


Figure 9. $\chi T/C$ against T for liquid ^{3}He; χ=susceptibility and C=Curie constant (*after* Fairbank, Ard and Walters 1954).

behaviour of an ideal Fermi–Dirac gas with the same density and atomic mass as liquid ^{3}He, i.e. with degeneracy temperature $T_0 \sim 5$°K. The experimental points are seen to deviate widely from the latter curve, a significant deviation from curve A only commencing at about 1·2°K. Curve B corresponds to a Fermi–Dirac gas with T_0=0·45°K, the degeneracy temperature which provides the best fit to the data. This agreement is probably fortuitous since the *thermal* properties of such an ideal F–D gas would be very different from those observed. Nevertheless one may conclude from the susceptibility data that the *nuclear-spin entropy* is considerably less than R ln 2 at 0·5°K. (For a more detailed discussion on the various experimental results and their possible theoretical interpretation† the reader is referred to the recent paper by Temperley (1954).)

† *Note added in proof.* See also L. Goldstein (*Phys. Rev.*, 1954, **96**, 1455), P. J. Price (*Ibid.*, 1955, **97**, 259) and O. K. Rice (*Ibid.*, 1955, **97**, 263).

§ 7. Further Developments in Magnetic Cooling

7.1. *Two-Stage Demagnetization*

To attain temperatures below 0·01°K from an initial temperature of 1°K it is usually necessary to have magnetic fields of about 25 kOe available. Such fields can only be produced by large iron-cored magnets or high-power solenoids, the installation of which is rather expensive. By carrying out a two stage demagnetization process, however, where the first stage is cooled to 0·1°K say, and is used as a sink for the heat of magnetization of the second stage, very much smaller fields can be used. Darby *et al.* (1951) describe such an experiment in which a field of only 9000 oersteds is used to obtain a final estimated temperature of 0·001°K. The first stage consisted of a compressed powder specimen of iron ammonium alum which was cooled to 0·25°K and the second stage a similar but smaller specimen of dilute potassium chromic alum.

The success of the experiment depended largely upon securing a very small heat leak into the second stage which had a small heat capacity, and upon the design of the thermal switch between the stages. The thermal switch, for which a superconducting wire was used, allows good thermal contact between the stages when the lower stage is being magnetized, but after demagnetization isolates them thermally (see § 5.1). The apparatus used is shown in figure 10 (*a*), and the warming up curve for the lower stage in figure 10 (*b*). Darby *et al.* actually used a lead wire

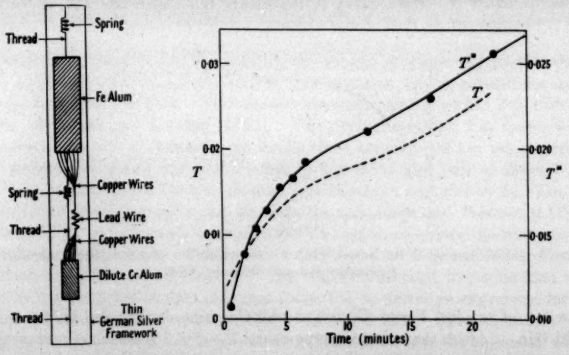


Figure 10. (*a*). Experimental arrangement for two-stage demagnetization. (*b*). Warm-up rate of lower stage (*after* Darby, Hatton, Rollin, Seymour and Silsbee 1951).

3 cm long and 0·3 mm diameter. The demagnetization of the upper stage and magnetization of the lower stage were carried out slowly so that heat transfer took place under conditions approaching reversibility.

7.2. *Nuclear Demagnetization*

The lower limit of temperature attainable by demagnetizing paramagnetic salts is probably about 0·001°K and is set by the strength of the interactions between the paramagnetic ions. Although the effect of the interactions can be reduced by

diluting the salt with suitable diamagnetic ions there is a limit to the usefulness of this technique since the heat capacity becomes very small, and at the limit of extreme dilution the dominant contribution to the entropy at 1°K will be that of the lattice. The suggestion was made independently by Gorter (1934) and Kurti and Simon (1935 d) that use could be made of nuclear paramagnetism to obtain still lower temperatures. The limit would be set by the strength of the direct interactions between nuclei but it is expected that temperatures as low as 10^{-6} deg K might be reached. The problem has been discussed in some detail by Simon (1939), who pointed out that if the temperatures reached after demagnetization are to be held for a usefully long period of time, even assuming a most favourable heat leak of 1 erg min^{-1}, initial magnetizing conditions of $H/T=10^7$ kilo-oersted deg^{-1} at least are needed. Since the biggest magnetic field which one supposes would be available for the purpose in the near future would be about 100 kilo-oersteds, a starting temperature of 0·01°K would be needed, and a two-stage demagnetization process would have to be carried out. Simon also lists substances which would be most suitable for the nuclear stage.

Although it is now twenty years since the suggestion was made, a successful experiment has not yet been carried out, although one should mention the interesting experiments of Rollin and Hatton (1949), Pound (1951) and Purcell and Pound (1951). The reason for this is that there are considerable difficulties in designing a switch which would be capable of transmitting the heat of magnetization from the nuclear stage to the electronic stage. Suppose that one has a large quantity of potassium chromic alum for the first stage which has been cooled somewhat below 0·01°K. In order to be able to reduce the stray field at the electronic stage to a sufficiently small value while maintaining an intense field at the nuclear stage, the two stages would have to be placed about 20 cm apart. The problem is to design the link so that the nuclear stage will be cooled to 0·01°K in a reasonable time, say an hour. (For the heat flow between the stages to be as nearly as possible reversible, and for ease of heat transfer within the first stage, it would be better to demagnetize the chrome alum initially to some intermediate temperature, say 0·1°K, and then slowly to reduce it to 0·01°K, but the present example will illustrate the difficulties involved.) The change in entropy dS due to a temperature change dT in a field H kilo-oersteds is given by (Simon 1939)

$$dS/dT=2{\cdot}15\times10^{-10}R\mu^2\frac{I+1}{I}\frac{2H^2}{T^3} \qquad \ldots\ldots(17)$$

where I is the spin and μ the nuclear magnetic moment in nuclear magnetons. From this formula and a knowledge of the thermal resistance controlling the heat flow, one can then compute how long the cooling of the nuclear stage will take.

The use of a copper rod of 1 cm^2 cross section has been shown by Kurti (1948) to provide adequate heat transfer. Two points must be borne in mind with this arrangement. First, copper has a large nuclear magnetic moment μ so that if it is situated in a very high magnetic field an appreciable amount of heat of magnetization will be produced which will also have to be conducted away; this problem could be avoided by having a short length of a more suitable material with small μ, such as lead, in the highest field; the lead could then also act as a thermal switch. Second, there may be appreciable eddy current heating due either to magnetic current ripple or even just through demagnetizing, so that the copper would have to be laminated.

Where the nuclear stage is a metal there would be no difficulty in making a good contact with the metallic link leading to the electronic stage. At the latter, however, the situation is quite different. For a mechanical type contact with a resistance equal to that given by equation (16) one finds that a contact area of the order of 10^5 cm² is needed. The situation if liquid helium is used is somewhat more favourable, provided one considers only the thermal conductivity (Kurti 1948). Assuming contact to be made through layers of thickness 0·01 cm then an area of 250 cm² will suffice. However, when one takes into account the surface resistance it turns out to be the most important factor and an area of the order of 50 000 cm² would be needed. One may summarize the difficulty in carrying out a nuclear demagnetization therefore as being that of overcoming contact resistance.

7.3. *Continuous Cooling Cycles*

The possibility of setting up a continuously operating refrigerator for producing temperatures below 1°K using a paramagnetic salt as working substance was pointed out by Daunt and Heer (1949 c). The advantage of such a system is that one would have a constant temperature bath below 1°K and so be in position to perform experiments in which quite big heat influxes could be tolerated. The steps of a cycle undergone by the working substance (ferric ammonium alum) working in a Carnot cycle are shown in figure 11. There are two isothermals AB and CD and two

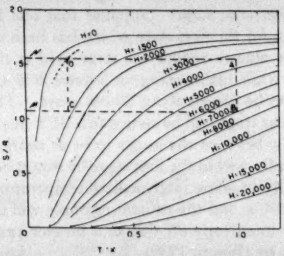


Figure 11. Entropy plotted against temperature for ferric ammonium alum for different magnetic fields, showing the ideal path for the working substance in a heat pump (*after* Heer, Barnes and Daunt 1954 b).

adiabatics BC and DA. Along AB the salt is magnetized isothermally at 1°K, the heat being rejected to a liquid helium bath. After thermal isolation it is demagnetized to the point C. It then takes in heat from the cold reservoir along CD, and at the same time the magnetic field is further reduced so that the temperature of the working substance remains constant. It is finally magnetized adiabatically along DA until its temperature reaches 1°K, when it is placed in thermal contact with the helium bath and the magnetization completed isothermally. The heat extracted from the cold reservoir $=T_C(S_D-S_C)$ (i.e. the area MNCD) where T_C is the temperature at C and S_C and S_D the entropies at C and D respectively. The material for the working substance is chosen so that this quantity is large, i.e. the paramagnetic ion should have a large value of J and should not deviate markedly from ideal behaviour in this region.

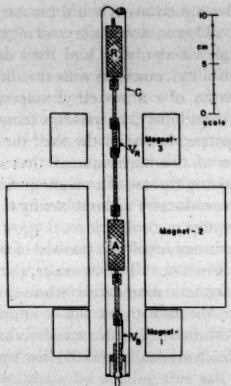


Figure 12. Schematic layout of a heat pump for producing and maintaining temperatures below 1°K (*after* Heer, Barnes and Daunt 1954 b).

Two experimental arrangements have been described, one by Collins and Zimmerman (1953) and the other by Heer, Barnes and Daunt (1953, 1954 a, b). In the former arrangement mechanical make-and-break thermal valves were used, which, however, caused an appreciable heating through the mechanical vibrations they introduced. A temperature of 0·73°K was maintained, the heat sink being at 1·13°K. In the latter arrangement, which is illustrated in figure 12, superconducting lead ' valves ' were used. The whole assembly is enclosed in a german silver tube C which is surrounded by liquid helium. The working substance A is 15 g of ferric ammonium alum and can be connected thermally either to the helium bath through valve V_B, or to the cold reservoir R, a 15 g sample of chromic potassium alum, through valve V_R. Magnets 1 and 3 are used for opening and closing the valves, and magnet 2, which can produce 7000 Oe, is used for magnetizing A.

The actual cycle deviated from ideality for several reasons. There were extraneous heat influxes, partly due to generation of eddy currents in the copper links in the system through switching the magnetic fields on and off, and partly due to an imperfect vacuum within C. The thermal contacts were imperfect and there was difficulty in following the true isothermal path CD.

Working on a 2-minute automatic cycling period, a reservoir temperature of 0·2°K could be reached. At 0·3°K the rate of heat extraction from the reservoir was $1{\cdot}4\times10^4$ ergs/cycle. Heer *et al.* pointed out that further developments of the thermal switches and contacts would bring about a more efficient refrigerator, and proposed addition of a further stage to enable temperatures of 0·05°K to be reached.

7.4. *The Use of Anisotropic Paramagnetic Salts*

It was proposed by Cooke (1953) that two novel types of experiment could be performed by making use of the fact that at very low temperatures certain rare earth salts (e.g. cerium ethylsulphate, cerium magnesium nitrate and neodymium magnesium nitrate) show paramagnetic susceptibilities which are extremely anisotropic.

In the first, adiabatic demagnetization could be carried out by rotation (Bogle, Cooke and Whitley 1951). That is, a single crystal of the salt could be magnetized at 1°K along an axis of high susceptibility and then demagnetized by rotating it until an axis of low susceptibility coincided with the direction of the applied field. For example, demagnetization of cerium ethyl sulphate from 0·95°K and 1800 oersteds along the high-susceptibility axis yields a temperature of 0·20°K. If the same field is then applied perpendicular to the axis, the temperature is raised only to 0·23°K. The advantage of this technique is that it is possible to carry out experiments below 1°K involving the use of a magnetic field (e.g. the investigation of the threshold field of a superconductor) without having to employ the more elaborate techniques of cooling by contact.

The second type of experiment involves a method of measuring the susceptibility of another (isotropic) salt below 1°K. The two salts, in thermal contact, are cooled together and the resultant magnetic moment determined in two different directions. Since the contribution from the anisotropic salt is known, that from the isotropic salt can be determined. Cerium magnesium nitrate is an ideal substance from this point of view, for not only is it markedly anisotropic, but it also obeys Curie's law down to 0·006°K.

§ 8. Nuclear Orientation

The production of oriented nuclear systems in the solid state has been a recent achievement of the magnetic cooling method and has proved to be of use in the study of problems in nuclear physics. We shall not attempt to discuss here, however, the nuclear data obtained. A detailed discussion of the four methods which have been proposed for producing oriented nuclei at low temperatures (two of which have been carried out so far) has been given by Simon, Rose and Jauch (1951) and by Bleaney (1951 a). A distinction is drawn between two types of orientation; one is a *polarization* where there is a net magnetic moment, and the other is an *alignment* where, although orientations close to a given axis or plane in the solid are preferred, there is no distinction between parallel and anti-parallel directions for the nuclear spins. The experiments which gave the first clear-cut demonstration of nuclear *alignment* were reported by Daniels, Grace and Robinson (1951) using a method proposed by Bleaney (1951 a); similar experiments were reported shortly afterwards by Gorter, Poppema, Steenland and Beun (1951). The first experiments in which a significant nuclear *polarization* was obtained were reported by Ambler, Grace, Halban, Kurti, Durand, Johnson and Lemmer (1953). In all these experiments ^{60}Co nuclei were oriented.

The most direct way of orienting nuclei, i.e. by applying a very large external magnetic field at a sufficiently low temperature, is essentially the same problem as that of nuclear demagnetization and a successful realization of the method is held up for the same technical reasons. However, the proposal made independently by Gorter (1948 b) and Rose (1949), that use could be made of the large magnetic field (10^5–10^6 oersteds) acting on nuclei of paramagnetic ions due to the electrons in the unfilled shell, has been successfully carried out. The interaction between this magnetic field and the nuclear magnetic moment in certain ions causes a large hyperfine splitting so that at temperatures of the order of 0·01°K only the lowest hyperfine level is populated, corresponding to an alignment of the nuclear spin with respect to the electron moment. Thus polarization of the electron moments, which can be achieved at these very low temperatures by a field of only a few hundred oersteds, results also in a polarization of the nuclear moments. In principle, the

procedure would be to magnetize a suitable salt at 1°K with a field strong enough to saturate the electron paramagnetism, then to demagnetize to a field of a few hundred oersteds. The method was tried by Gorter, de Klerk, Poppema, Steenland and de Vries (1949) with ferric ammonium alum, but only a very small and non-reproducible effect was produced. Later it was effectively used by Bernstein, Roberts, Stanford, Dabbs and Stephenson (1954) to produce a polarization of about 17% of the saturation value of ^{55}Mn nuclei in manganous ammonium sulphate. It has been pointed out, however, that this method cannot be employed to produce very high degrees of nuclear polarization since the hyperfine splitting limits the lowest temperature which can be attained. We can see that magnetization at 1°K will remove most of the entropy from the electron spin system but practically none from the nuclear spin system, whereas for high degrees of nuclear polarization we must remove as much entropy as possible from both systems. It was suggested by Gorter (1948 a) that the salt under investigation could be cooled by contact with another salt having no hyperfine structure. Although in this case one encounters the difficulty of contact cooling below 1°K, the method has been used by Roberts, Bernstein, Dabbs and Stanford (1954) to produce a 12% polarization of ^{149}Sm nuclei in samarium ethyl sulphate by cooling this salt to 0·12°K by contact with ferric ammonium alum at 0·06°K. A more direct method is to incorporate in the same salt ions which have a very small hyperfine structure or none at all (see for example Daniels 1951). Yet another consideration was put forward by Bleaney (1951 b), namely that in the salt chosen the small externally applied magnetic field should be the dominant force tending to align the electron moments. In many salts this is not the case and there are strong forces arising from the effect of the crystalline electric field, unless a salt is chosen in which the ions are magnetically isotropic. Both of the above conditions were fulfilled in the experiment by Ambler *et al.* (1953) in which over a 90% polarization of the cobalt nuclei, incorporated in the salt $3[0{\cdot}5\%\ \mathrm{Co},\ 99{\cdot}5\%\ \mathrm{Mg}\ (\mathrm{NO_3})_2]\ 2\mathrm{Ce}(\mathrm{NO_3})_3.24\mathrm{H_2O}$, was obtained.

In the experiment of Daniels, Grace and Robinson (1951) ^{60}Co nuclei were aligned in a Tutton salt of composition $(1\%\mathrm{Co}, 12\%\mathrm{Cu}, 87\%\mathrm{Zn})\mathrm{Rb_2}(\mathrm{SO_4})_2.6\mathrm{H_2O}$. The method relies on the fact that in certain paramagnetic salts, in particular the cobalt Tutton salts, magnetic interaction between the unpaired electrons of the magnetic ion and the nucleus is anisotropic due to the effect of the crystalline electric field; the hyperfine splitting then depends not only upon the relative orientations of the electronic and nuclear spins, but also upon their orientation with respect to the axis of symmetry of the crystalline electric field. In other words, the magnitude of the magnetic field at the nucleus depends upon the orientation of the electron spin with respect to the symmetry axis. Alignment occurs, at a sufficiently low temperature, in zero external field; we have an alignment rather than polarization because the energy of an ion is unaltered by reversing the directions of both nuclear and electron spins. Data on the suitability of salts for this type of alignment can be readily obtained by paramagnetic resonance and have been discussed by Bleaney (1951 a, b) in terms of the constants D, A and B appearing in the spin-Hamiltonian.

The fourth method of orienting nuclei, suggested by Pound (1949), produces an alignment and relies on the interaction between the nuclear quadrupole moment and an electric field gradient. A suitable substance would be one in which the associated hyperfine splitting is large and anisotropic. The difficulty of finding such a

substance and cooling it to a sufficiently low temperature has so far precluded a successful realization of this suggestion.

The experiments which can be carried out with oriented nuclei fall into two groups. One group, for which polarized nuclei must be used, comprises a study of the scattering of polarized nucleons by polarized nuclei. The other, for which either polarized or aligned nuclei may be used, comprises a study of the directional effects in emission from radioactive nuclei. The experiments which have been performed in the former class to date have been carried out at the Oak Ridge National Laboratory, U.S.A., by Roberts, Bernstein and collaborators, on the interaction of polarized slow neutrons with polarized ^{55}Mn and ^{149}Sm nuclei. When polarized slow neutrons are captured or scattered by a target in which the nuclear spins are also polarized, the cross section is different from that observed when the spins are randomly oriented.

In the ^{55}Mn (Bernstein *et al.* 1954, Dabbs and Roberts 1954) experiment, the change in cross section was determined by measuring the difference between the numbers of radioactive ^{56}Mn nuclei produced. The 2·6-hour half-life of ^{55}Mn was long enough to permit the counting to be done after the sample had been removed from the cryostat.

In the ^{149}Sm experiment a somewhat different method was used, the nuclear polarization being measured by determining the state of polarization of an originally unpolarized beam of neutrons after transmission through a samarium salt. The salt actually used, samarium ethyl sulphate, is unsuitable for self-cooling, and ferric ammonium alum was employed as a coolant, thermal contact being established by means of a copper rod.

The experiments in the second group comprise a study of the emission pattern of γ-rays from aligned radioactive nuclei (Spiers 1948). The polar diagram for the intensity can be written

$$I(\theta)=\sum_M W_M I_M(\theta) \qquad \ldots\ldots(18)$$

where W_M is the relative population of the nuclear magnetic substate M and $I_M(\theta)$ is the angular distribution for the radiation originating from this state (θ being the angle between the axis of alignment and direction of emission). $I_M(\theta)$ depends only upon the spin and spin changes involved in the decay (see for example Tolhoek and Cox 1953, Steenberg 1953). It is a polynomial in $\cos^2\theta$ containing terms up to $\cos^{2l}\theta$ where l is the multipolarity of the emitted γ-ray. W_M depends only on the hyperfine structure and the temperature. A measurement of the angular distribution can be used to obtain, through the term $I_M(\theta)$ in equation (18), information about the decay scheme of the aligned nucleus; it is especially easy to obtain the multipolarity of a γ-ray transition. Through the term W_M in equation (18) one can also obtain information concerning the hyperfine splitting, and hence a value for the nuclear magnetic moment.

Many experiments of this type have been reported, but since the conclusions lie in the realm of nuclear physics they will only be mentioned here and not described in detail: ^{60}Co by Bleaney *et al.* (1952), Bleaney *et al.* (1954), Poppema *et al.* (1952), Gorter *et al.* (1952), Bishop *et al.* (1952); ^{58}Co by Daniels *et al.* (1952), Grace and Halban (1952), Bishop *et al.* (1952); ^{54}Mn by Daniels and Kurti (1953); Grace *et al.* (1954) and Bishop *et al.* (1954); and ^{141}Ce and ^{147}Nd by Ambler, Hudson and Temmer (1955).

Acknowledgments

The writers wish to express their gratitude to the various authors and editors of journals for permission to reproduce the figures used in this report; to Dr. A. H. Cooke and Dr. N. Kurti for providing the data for tables 1 and 3; and to Dr. J. G. Daunt, Dr. W. M. Fairbank and Dr. L. D. Roberts for making available manuscripts in advance of publication. Finally it is a great pleasure to record their indebtedness to Dr. N. Kurti for many helpful criticisms.

References

Abragam, A., and Pryce, M. H. L., 1951, *Proc. Roy. Soc.* A, **205,** 135.
Allen, J. F., and Mendoza, E., 1947, *Proc. Camb. Phil. Soc.*, **44,** 280.
Allen, J. F., and Shire, E. S., 1937, *Nature, Lond.*, **139,** 878.
Ambler, E., 1953, *Thesis*, University of Oxford.
Ambler, E., Grace, M. A., Halban, H., Kurti, N., Durand, H., Johnson, C. E., and Lemmer, H. R., 1953, *Phil. Mag.*, **44,** 216.
Ambler, E., and Hudson, R. P., 1954 a, *Phys. Rev.*, **95,** 1143; 1954 b, *Ibid.*, **96,** 907.
Ambler, E., Hudson, R. P., and Temmer, G. M., 1955, *Phys. Rev.*, **97,** 1212.
Ambler, E., and Kurti, N., 1952 a, *Phil. Mag.*, **43,** 1307; 1952 b, *Ibid.*, **43,** 260.
Andrews, F. A., Spohr, D. A., and Webber, R. T., 1950, *Phys. Rev.*, **78,** 342.
Ashmead, J., 1944, *Thesis*, University of Cambridge.
Atkins, K. R., 1953, *Phys. Rev.*, **89,** 526.
Atkins, K. R., and Osborne, D. V., 1950, *Phil. Mag.*, **41,** 1078.
Beenakker, J. J. M., Taconis, K. W., Lynton, E. A., Dokoupil, Z., and Van Soest, G., 1952, *Physica*, **18,** 433.
Benzie, R. J., and Cooke, A. H., 1949, *Nature, Lond.*, **164,** 837; 1950 a, *Proc. Phys. Soc.* A, **63,** 201; 1950 b, *Ibid.*, **63,** 213.
Berman, R., 1951, *Proc. Roy. Soc.* A, **208,** 90; 1953, *Advances in Physics* (*Phil. Mag.* Suppl.), **2,** 103.
Berman, R., and MacDonald, D. K. C., 1952, *Proc. Roy. Soc.* A, **211,** 122.
Bernstein, S., Roberts, L. D., Stanford, C. P., Dabbs, J. W. T., and Stephenson, T. E., 1954, *Phys. Rev.*, **94,** 1243.
Bethe, H. A., 1929, *Ann. Phys., Lpz.*, **3,** 133.
Bijl, D., 1949, *Physica*, **14,** 684.
Bishop, G. R., Daniels, J. M., Durand, H., Johnson, C. E., and Perez, J., 1954, *Phil. Mag.*, **45,** 1197.
Bishop, G. R., Daniels, J. M., Goldschmidt, G., Halban, H., Kurti, N., and Robinson, F. N. H., 1952, *Phys. Rev.*, **88,** 1432.
Bitter, F., 1936, *Rev. Sci. Instrum.*, **7,** 479, 483.
Blaisse, B. S., Cooke, A. H., and Hull, R. A., 1939, *Physica*, **6,** 231.
Bleaney, B., 1950 a, *Phys. Rev.*, **78,** 214; 1950 b, *Proc. Roy. Soc.* A, **204,** 216; 1950 c, *Ibid.*, **204,** 203; 1951 a, *Phil. Mag.*, **42,** 441; 1951 b, *Proc. Phys. Soc.* A, **64,** 315.
Bleaney, B., and Bowers, K. D., 1951, *Proc. Phys. Soc.* A, **64,** 1135.
Bleaney, B., Bowers, K. D., and Ingram, D. J. E., 1951, *Proc. Phys. Soc.* A, **64,** 758.
Bleaney, B., Daniels, J. M., Grace, M. A., Halban, H., Kurti, N., and Robinson, F. N. H., 1952, *Phys. Rev.*, **85,** 688.
Bleaney, B., Daniels, J. M., Grace, M. A., Halban, H., Kurti, N., Robinson, F. N. H., and Simon, F. E., 1954, *Proc. Roy. Soc.* A, **221,** 170.
Bleaney, B., Elliott, R. J., Scovil, H. E. D., and Trenam, R. S., 1951, *Phil. Mag.*, **42,** 1062.
Bleaney, B., and Ingram, D. J. E., 1951 a, *Proc. Roy. Soc.* A, **205,** 336; 1951 b, *Ibid.*, **208,** 143.
Bleaney, B., and Stevens, K. W. H., 1953, *Rep. Progr. Phys.*, **16,** 108 (London: Physical Society).
Bogle, G. S., Cooke, A. H., and Whitley, S., 1951 *Proc. Phys. Soc.* A, **64,** 931.
Bogle, G. S., and Heine V., 1954, *Proc. Phys. Soc.* A, **67,** 734.
Casimir, H. B. G., and de Haas, W. J., 1940, *Physica*, **7,** 70.

CASIMIR, H. B. G., DE HAAS, W. J., and DE KLERK, D., 1939 a, *Physica*, **6**, 241; 1939 b, *Ibid.*, **6**, 255.

CASIMIR, H. B. G., DE KLERK, D., and POLDER, D., 1940, *Physica*, **7**, 737.

CLEMENT, J. R., QUINNELL, E. H., STEELE, M. C., HEIN, R. A., and DOLECEK, R. L., 1953, *Rev. Sci. Instrum.*, **24**, 545.

COCKCROFT, J. D., 1928, *Phil. Trans. Roy. Soc.*, **227**, 317.

COLLINS, S. C., 1950, M.I.T. Technical Rep. to O.N.R. No. 1 (Contract No. N5 ori-78) —unpublished.

COLLINS, S. C., and ZIMMERMAN, J., 1953, *Phys. Rev.*, **90**, 991.

COOKE, A. H., 1949, *Proc. Phys. Soc.* A, **62**, 269; 1950, *Rep. Progr. Phys.* (London: Physical Society), **13**, 276; 1953, *Third International Conference On Low Temperature Physics and Chemistry*; Rice Institute, Houston, Texas. Abstracts, p. 27 (unpublished).

COOKE, A. H., DUFFUS, H. J., and WOLF, W. P., 1953, *Phil. Mag.*, **44**, 623.

COOKE, A. H., and HULL, R. A., 1937, *Proc. Roy. Soc.* A, **162**, 404; 1939, *Nature, Lond.*, **143**, 799; 1942, *Proc. Roy. Soc.* A, **181**, 83.

CROFT, A. J., FAULKNER, E. A., HATTON, J. and SEYMOUR, E. F. W., 1953, *Phil. Mag.*, **44**, 289.

DABBS, J. W. T., and ROBERTS, L. D., 1954, *Phys. Rev.*, **95**, 970.

DANIELS, J. M., 1950, *Proc. Phys. Soc.* B, 63, 1028; 1951, *Proc. International Conference on Low Temperature Physics*, Oxford 1951, p. 159 (unpublished); 1953 a, *Proc. Phys. Soc.* A, **66**, 673; 1953 b, *Brit. J. Appl. Phys.*, **4**, 50.

DANIELS, J. M., GRACE, M. A., HALBAN, H., KURTI, N. and ROBINSON, F. N. H., 1952, *Phil. Mag.*, **43**, 1297.

DANIELS, J. M., GRACE, M. A. and ROBINSON, F. N. H., 1951, *Nature, Lond.*, **168**, 780.

DANIELS, J. M., and KURTI, N., 1953, *Third International Conference on Low Temperature Physics and Chemistry*; Rice Institute, Houston, Texas. Abstracts, p. 25 (unpublished); 1954, *Proc. Roy. Soc.* A, **221**, 243.

DANIELS, J. M., and ROBINSON, F. N. H., 1953, *Phil. Mag.*, **44**, 630.

DARBY, J., HATTON, J., ROLLIN, B. V., SEYMOUR, E. F. W. and SILSBEE, H. B., 1951, *Proc. Phys. Soc.* A, **64**, 861.

DAUNT, J. G., and HEER, C. V., 1949 a, *Phys. Rev.*, **76**, 715; 1949 b, *Ibid.*, **76**, 1324; 1949 c, *Ibid.*, **76**, 985; 1950, *Ibid.*, **79**, 46.

DEBYE, P., 1914, *Vorträge über die kinetische Theorie der Materie und der Elektrizität* (Leipzig: Teubner); 1926, *Ann. Phys., Lpz.*, **81**, 1154.

FAIRBANK, H. A., and WILKS, J., 1954, *Phys. Rev.*, **95**, 277.

FAIRBANK, W. M., ARD, W. B., and WALTERS, G. K., 1954, *Phys. Rev.*, **95**, 566.

GARDNER, W. E., and KURTI, N., 1954, *Proc. Roy. Soc.* A, **223**, 542.

GARRETT, C. G. B., 1950 a, *Proc. Roy. Soc.* A, **203**, 375; 1950 b, *Ibid.*, **203**, 392; 1950 c, *Phil. Mag.*, **41**, 621; 1951 a, *Proc. Roy. Soc.* A, **206**, 242; 1951 b, *J. Chem. Phys.*, **19**, 1154.

GAUME, F., 1946, *C. R. Acad. Sci., Paris*, **223**, 719.

GERRITSEN, A. N., and KORRINGA, J., 1951, *Phys. Rev.*, **84**, 604.

GERRITSEN, H. J., OKKES, R., GIJSMAN, H. M., and VAN DEN HANDEL, J., 1954, *Physica*, **20**, 13.

GIAUQUE, W. F., 1927, *J. Amer. Chem. Soc.*, **49**, 1864; 1936, *Industr. Engng. Chem.*, **28**, 743; 1953, *Phys. Rev.*, **92**, 1339.

GIAUQUE, W. F., and MACDOUGALL, D. P., 1933, *Phys. Rev.*, **43**, 768; 1935 a, *Ibid.*, **47**, 885; 1935 b, *J. Amer. Chem. Soc.*, **57**, 1175; 1938, *Ibid.*, **60**, 375.

GIAUQUE, W. F., STOUT, J. W., and CLARK, C. W., 1938, *J. Amer. Chem. Soc.*, **60**, 1053.

GIAUQUE, W. F., STOUT, J. W., EGAN, C. J., and CLARK, C. W., 1941, *J. Amer. Chem. Soc.*, **63**, 405.

GOODMAN, B. B., 1951, *Nature, Lond.*, **167**, 111; 1953, *Proc. Phys. Soc.* A, **66**, 217.

GOODMAN, B. B., and MENDOZA, E., 1951, *Phil. Mag.*, **42**, 594.

GORTER, C. J., 1934, *Phys. Z.*, **35**, 923; 1948 a, *Les Phénomènes Cryomagnétiques*, Langevin-Perrin Colloquium, Collège de France, p. 76; 1948 b, *Physica*, **14**, 504.

GORTER, C. J., and HAANTJES, J., 1952, *Physica*, **18**, 285.

GORTER, C. J., DE KLERK, D., POPPEMA, O., STEENLAND, M. J., and DE VRIES, H., 1949, *Physica*, **15**, 679.

GORTER, C. J., POPPEMA, O., STEENLAND, M. J., and BEUN, J. A., 1951, *Physica*, **17**, 1050.

GORTER, C. J., TACONIS, K. W., and BEENAKKER, J. J. M., 1951, *Physica*, **17**, 841.

GORTER, C. J., TOLHOEK, H. A., POPPEMA, O., STEENLAND, M. J., and BEUN, J. A., 1952, *Physica*, **18,** 135.
GRACE, M. A., and HALBAN, H., 1952, *Physica*, **18,** 1227.
GRACE, M. A., JOHNSON, C. E., KURTI, N., LEMMER, H. R., and ROBINSON, F. N. H., 1954, *Phil. Mag.* **45,** 1192.
DE HAAS, W. J., CASIMIR, H. B. G., and VAN DEN BERG, G. J., 1938, *Physica*, **5,** 225.
DE HAAS, W. J., and WIERSMA, E. C., 1933, *Physica*, **1,** 779; 1936, *Ibid.*, **3,** 491.
DE HAAS, W. J., WIERSMA, E. C., and KRAMERS, H. A., 1933 a, *Nature, Lond.*, **131,** 719 and *C.R. Acad. Sci., Paris*, **196,** 1975; 1933 b, *Physica*, **1,** 1.
VAN DEN HANDEL, J., 1952, *Physica*, **18,** 921.
HEBB, M. H., and PURCELL, E. M., 1937, *J. Chem. Phys.*, **5,** 338.
HEER, C. V., BARNES, C. B., and DAUNT, J. G., 1953, *Phys. Rev.*, **91,** 412; 1954 a, *Ibid.*, **93,** 362; 1954 b, *Rev. Sci. Instrum.*, **25,** 1088.
HEER, C. V., and DAUNT, J. G., 1949, *Phys. Rev.*, **76,** 854.
HEER, C. V., and RAUSCH, C., 1953, *Phys. Rev.*, **90,** 530.
HEISENBERG, W., 1948, *Z. Naturforsch.*, **3a,** 65.
HOLDEN, A. N., KITTEL, C., MERRITT, F. R., and YAGER, W. A., 1950, *Phys. Rev.*, **77,** 147.
HOLDEN, A. N., YAGER, W. A., and MERRITT, F. R., 1951, *J. Chem. Phys.*, **19,** 1319.
HOWLING, D. H., DARNELL, F. J., and MENDOZA, E., 1954, *Phys. Rev.*, **93,** 1416.
HUDSON, R. P., 1949 a, *J. Sci. Instrum.*, **26,** 401; 1949 b, *Thesis*, University of Oxford; 1952, *Phys. Rev.*, **88,** 570.
HUDSON, R. P., HUNT, B., and KURTI, N., 1949, *Proc. Phys. Soc.* A, **62,** 392.
HUDSON, R. P., and MCLANE, C. K., 1954, *Phys. Rev.*, **95,** 932.
HULL, J. R., and HULL, R. A., 1941, *J. Chem. Phys.*, **9,** 465.
HULL, R. A., 1947, *Report of Cambridge Conference on Low Temperatures*, p. 72 (London: Physical Society).
HULL, R. A., WILKINSON, K. R., and WILKS, J., 1951, *Proc. Phys. Soc.* A, **64,** 379.
HUTCHINSON, C. A., 1949, *Phys. Rev.*, **75,** 1769.
KAPITZA, P. L., 1941, *J. Phys., Moscow*, **4,** 181.
KEESOM, P. H., 1948, *Thesis*, University of Leiden.
KEESOM, W. H., and WESTMIJZE, W. K., 1941, *Physica*, **8,** 1044.
KHALATNIKOV, I. M., 1952, *J. Exp. Theor. Phys.*, **22,** 687.
KING, J. C., and FAIRBANK, H. A., 1953, *Phys. Rev.*, **90,** 989.
DE KLERK, D., 1946, *Physica*, **12,** 513; 1948, *Thesis*, University of Leiden; 1952, *U.S. Nat. Bur. Stds. Circular No.* 519.
DE KLERK, D., and HUDSON, R. P., 1953, *Phys. Rev.*, **91,** 278; 1954, *J. Res. Nat. Bur. Stand.*, **53,** 173.
DE KLERK, D., HUDSON, R. P., and PELLAM, J. R., 1953, *Phys. Rev.*, **89,** 326; 1954, *Ibid.*, **93,** 28.
DE KLERK, D., STEENLAND, M. J., and GORTER, C. J., 1948, *Nature, Lond.*, **161,** 678; 1949, *Physica*, **15,** 649.
KRAMERS, H. A., 1930, *Proc. Acad. Sci., Amst.*, **33,** 953; 1934, *Physica*, **1,** 182.
KRAMERS, H. C., WASSCHER, J. D., and GORTER, C. J., 1952, *Physica*, **18,** 329.
KRAUSS, D. L., and NUTTING, G. C., 1941, *J. Chem. Phys.*, **9,** 133.
KRONIG, R., and THELLUNG, A., 1950, *Physica*, **16,** 678.
KURTI, N., 1933, *Z. Phys. Chem.* B, **20,** 305; 1948, *Les Phénomènes Cryomagnétiques*, Langevin-Perrin Colloquium, Collège de France, p. 29; 1951, *J. Phys. Radium*, **12,** 282; 1952, *Low Temperature Physics* (London: Pergamon Press), p. 44.
KURTI, N., LAINÉ, P., ROLLIN, B. V., and SIMON, F. E., 1936 a, *C.R. Acad. Sci., Paris*, **202,** 1576; 1936 b, *Ibid.*, **202,** 1421.
KURTI, N., LAINÉ, P., and SIMON, F. E., 1937, *C.R. Acad. Sci., Paris*, **204,** 675.
KURTI, N., ROLLIN, B. V., and SIMON, F. E., 1936, *Physica*, **3,** 266.
KURTI, N., and SIMON, F. E., 1934, *Nature, Lond.*, **133,** 907; 1935 a, *Ibid.*, **135,** 31; 1935 b, *Proc. Roy. Soc.* A, **151,** 610; 1935 c, *Ibid.* **152,** 1; 1935 d, *Ibid.* **149,** 152; 1938 a, *Nature, Lond.*, **207,** 142; 1938 b, *Phil. Mag.*, **26,** 849; 1938 c, *Ibid.*, **26,** 840; 1953, *Ibid.*, **44,** 501.
LANDAU, L., 1941, *J. Phys., Moscow*, **5,** 71; 1947, *Ibid.*, **11,** 91.
LANGEVIN, P., 1905, *Ann. Chim.* (*Phys.*), **5,** 70.

LASAREW, B. G., and ESSELSON, B. N., 1941, *J. Phys., Moscow*, **5**, 151.
LLOYD, J. P., and PAKE, G. E., 1954, *Phys. Rev.*, **94**, 579.
MACDONALD, D. K. C., 1951, *Phil. Mag.*, **42**, 756; 1952, *Phys. Rev.*, **88**, 148.
MACDOUGALL, D. P., and GIAUQUE, W. F., 1936, *J. Amer. Chem. Soc.*, **58**, 1032.
MALAKER, S. F., 1951, *Phys. Rev.*, **84**, 133.
MAYPER, V., and HERLIN, M. A., 1953, *Phys. Rev.*, **89**, 523.
MEIJER, P. H. E., 1951, *Physica*, **17**, 899.
MENDELSSOHN, K., and OLSEN, J. L., 1950 a, *Proc. Phys. Soc.* A, **63**, 1182; 1950 b, *Phys. Rev.*, **80**, 859.
MENDELSSOHN, K., and RENTON, C. A., 1953, *Phil. Mag.*, **44**, 776.
MENDELSSOHN, K., and ROSENBERG, H. M., 1952, *Proc. Phys. Soc.* A, **65**, 385.
MENDOZA, E., 1948, *Les Phénomènes Cryomagnétiques*, Langevin-Perrin Colloquium, Collège de France, p. 53.
MENDOZA, E., and THOMAS, J. G., 1951, *Phil. Mag.*, **42**, 291.
NICOL, J., and DAUNT, J. G., 1952, *Proc. Cryogenics Conference, Schenectady*, p. 68.
NICOL, J., and TSENG, T. P., 1953, *Phys. Rev.*, **92**, 1062.
OLSEN, J. L., and RENTON, C. A., 1952, *Phil. Mag.*, **43**, 946.
ONSAGER, L., 1936, *J. Amer. Chem. Soc.*, **58**, 1486.
OPECHOWSKI, W., 1948, *Physica*, **14**, 237.
OSBORNE, D. V., 1951, *Proc. Phys. Soc.* A, **64**, 114.
OSBORNE, D. W., ABRAHAM, B. M., and WEINSTOCK, B., 1954, *Phys. Rev.*, **94**, 202.
PAKE, G. E., TOWNSEND, J., and WEISSMAN, S. I., 1952, *Phys. Rev.*, **85**, 682.
PEIERLS, R. E., 1929, *Ann. Phys., Lpz.*, **3**, 1055.
PELLAM, J. R., and SCOTT, R. B., 1949, *Phys. Rev.*, **76**, 869.
PENNEY, W. G., and SCHLAPP, R., 1932, *Phys. Rev.*, **41**, 194.
PICKARD, G., and SIMON, F. E., 1939, *Abstr. Commun. Roy. Soc. London* 521 (Apr. 3, 1939).
PIRENNE, J., 1949, *Helv. Phys. Acta*, **22**, 479.
PLATZMAN, R. L., 1953, *Phil. Mag.*, **44**, 497.
POMERANCHUK, I., 1949, *J. Exp. Theor. Phys.*, **19**, 42; 1950, *Ibid.*, **20**, 919.
POPPEMA, O., BEUN, J. A., STEENLAND, M. J., and GORTER, C. J., 1952, *Physica*, **18**, 1235.
POUND, R. V., 1949, *Phys. Rev.*, **76**, 1410; 1951, *Ibid.*, **81**, 156.
PURCELL, E. M., and POUND, R. V., 1951, *Phys. Rev.*, **81**, 279.
ROBERTS, L. D., BERNSTEIN, S., DABBS, J. W. T., and STANFORD, C. P., 1954, *Phys. Rev.*, **95**, 105.
ROBERTS, T. R., and SYDORIAK, S. G., 1954, *Phys. Rev.*, **93**, 1418.
ROLLIN, B. V., 1936, *Rep. VIIe Congrès Intern. du Froid*, **1**, 187.
ROLLIN, B. V., and HATTON, J., 1949, *Proc. Roy. Soc.* A, **199**, 237.
ROLLIN, B. V., and SIMON, F. E., 1939, *Physica*, **6**, 219.
ROSE, M. E., 1949, *Phys. Rev.*, **75**, 213.
SAMOILOV, B. N., 1951, *C.R. Acad. Sci., U.R.S.S.*, **81**, 791; 1952, *Ibid.*, **86**, 281.
SAUER, J., and TEMPERLEY, H. N. V., 1940, *Proc. Roy. Soc.* A, **176**, 203.
SCHMID, L. P., and SMART, J. S., 1954, *U.S. Naval Ordnance Rep. No.* 3640.
SCHNEIDER, E. E., and ENGLAND, T. S., 1951, *Physica*, **17**, 221.
SHIRE, E. S., and ALLEN, J. F., 1938, *Proc. Camb. Phil. Soc.*, **34**, 301.
SHIRE, E. S., and BARKLA, H. M., 1939, *Proc. Camb. Phil. Soc.*, **35**, 327.
SIMON, A., ROSE, M. E., and JAUCH, M. J., 1951, *Phys. Rev.*, **84**, 1155.
SIMON, F. E., 1935, *Nature, Lond.*, **135**, 763; 1939, *C.R. Congr. sur le Magnétisme*, p. 1, Strasbourg.
SLATER, J. C., 1951, *Phys. Rev.*, **84**, 179.
SMITH, T. S., and DAUNT, J. G., 1952, *Phys. Rev.*, **88**, 1172.
SMITH, T. S., GAGER, W. B., and DAUNT, J. G., 1953, *Phys. Rev.*, **89**, 654.
SPIERS, J. A., 1948, *Nature, Lond.*, **161**, 807.
STEELE, M. C., and HEIN, R. A., 1952, *Phys. Rev.*, **87**, 908.
STEENBERG, N. R., 1953, *Proc. Phys. Soc.* A, **66**, 399.
STEENLAND, M. J., 1952, *Thesis*, University of Leiden.
STEENLAND, M. J., DE KLERK, D., BEUN, J. A., and GORTER, C. J., 1951, *Physica*, **17**, 161.
STEENLAND, M. J., DE KLERK, D., and GORTER, C. J., 1949, *Physica*, **15**, 711.
STEENLAND, M. J., DE KLERK, D., POTTERS, M. L., and GORTER, C. J., 1951, *Physica*, **17**, 149.

STEENLAND, M. J., VAN DER MAREL, L. C., DE KLERK, D., and GORTER, C. J., 1949, *Physica*, **15**, 906.
TCHEN, T. C., and LONDON, F., 1953, *Phys. Rev.*, **89**, 1038.
TEMPERLEY, H. N. V., 1954, *Proc. Phys. Soc.* A, **67**, 495.
THOMAS, J. G., and MENDOZA, E., 1952, *Phil. Mag.*, **43**, 900.
TISZA, L., 1940, *J. Phys. Radium*, **1**, 165 and 350.
TOLHOEK, H. A., and COX, J. A. M., 1953, *Physica*, **19**, 101.
TRENAM, R. S., 1953, *Proc. Phys. Soc.* A, **66**, 118.
TSAI, B., 1947, *Report of Cambridge Conference on Low Temperatures*, p. 89 (London: Physical Society).
UBBINK, J., POULIS, J. A., and GORTER, C. J., 1951, *Physica*, **17**, 213.
VAN DIJK, H., 1946, *Physica*, **12**, 371.
VAN DIJK, H., and KEESOM, W. H., 1940, *Physica*, **7**, 970.
VAN DIJK, H., KEESOM, W. H., and STELLER, J. P., 1938, *Physica*, **5**, 625.
VAN VLECK, J. H., 1932, *Theory of Electric and Magnetic Susceptibilities* (Oxford: University Press), Ch. vii; 1937, *J. Chem. Phys.*, **5**, 320.
VAN VLECK, J. H., and PENNEY, W. G., 1934, *Phil. Mag.*, **17**, 961.
DE VRIES, G., and DAUNT, J. G., 1954, *Phys. Rev.*, **93**, 631.
WALLER, I., 1936, *Z. Phys.*, **104**, 132.
WEBB, F. J., WILKINSON, K. R., and WILKS, J., 1952, *Proc. Roy. Soc.* A, **214**, 546.
WEINSTOCK, B., ABRAHAM, B. M., and OSBORNE, D. W., 1952, *Phys. Rev.*, **85**, 158; 1953, *Ibid.*, **89**, 787.
WHITMER, C. A., and WEIDNER, R. T., 1951, *Phys. Rev.*, **84**, 159.
WILKINSON, K. R., and WILKS, J., 1951, *Proc. Phys. Soc.* A, **64**, 89.
WOLTJER, H. R., and KAMERLINGH ONNES, H., 1923, *Proc. Acad. Sci., Amst.*, **32**, 772 (Leiden Commun. 167c).
ZIMAN, J. M., 1954, *Phil. Mag.*, **45**, 100.

Attention is drawn to the following more general references which were not referred to in the text:

ALLEN, J. F., KURTI, N., MENDELSSOHN, K., and SIMON, F. E., 1952, *Low Temperature Physics: Four Lectures* (London: Pergamon Press).
CASIMIR, H.B.G., 1940, *Magnetism and Very Low Temperatures* (Cambridge: University Press).
GARRETT, C. G. B., 1954, *Magnetic Cooling* (Harvard: University Press).
GORDY, W., SMITH, W. V., and TRAMBARULO, R. F., 1953, *Microwave Spectroscopy* (New York: John Wiley).
GORTER, C. J., *Paramagnetic Resonance* (Amsterdam: Elsevier).
SQUIRE, C., 1953, *Low Temperature Physics* (New York: McGraw-Hill).
VAN VLECK, J. H., 1939, *Ann. Inst. Poincaré*, **10**, 57.

PARAMAGNETIC RESONANCE II

By K. D. BOWERS and J. OWEN

The Clarendon Laboratory, Oxford

CONTENTS

PAGE

§ 1. Introduction 305
§ 2. The paramagnetism of transition group ions in crystals 306
2.1. The 3d, 4d and 5d groups (octahedral complexes) 308
2.2. The 4f group 312
2.3. The 5f–(6d) group 313
§ 3. The lowest energy levels and the paramagnetic resonance spectrum .. 313
3.1. The resonance phenomenon 314
3.2. The spin Hamiltonian 314
3.3. Derivation of energy levels from spin Hamiltonian 319
§ 4. Crystallography 322
4.1. The octahedral complex 322
4.2. The axes 323
4.3. Crystal structures 324
§ 5. Experimental data 329
5.1. The spin Hamiltonian 330
5.2. Key to formulae and data 331
5.3. The 3d group 333
5.4. The 4d and 5d groups 355
5.5. The 4f group 362
5.6. The 5f–(6d) group 369
Acknowledgments 370
References 371

Abstract. In this report, which is complementary to that of Bleaney and Stevens in *Reports on Progress in Physics*, Volume XVI, paramagnetic resonance data are collected together for crystalline solids containing ions of the transition groups, and those parts of the theory necessary for an understanding of the results are presented in a fairly simple way. After a brief introduction, the magnetic properties of various types of compounds are considered, and a simple discussion is given of the spin Hamiltonian which is used to describe the lowest energy levels. Some details are also given of the structures of the ten crystal types which are most commonly used, and of the relation between the structure and the resonance spectrum of a single crystal. Finally, the resonance data which are available for ions of the 3d, 4d, 5d, 4f and 5f transition groups are tabulated. These data are expressed as values of the parameters in the appropriate spin Hamiltonian, and, where applicable, formulae are given for (1) energy levels, (2) susceptibility, (3) specific heat tail CT^2/R, in terms of these parameters.

§ 1. Introduction

During the past few years the paramagnetic resonance method has been extensively applied to investigate the lowest energy levels of paramagnetic ions in crystals. In principle this method is very simple. A steady magnetic field is applied to the paramagnetic ions, and the electronic energy levels undergo Zeeman splittings. Simultaneously, a small radio-frequency magnetic field is also applied, and the splittings can be measured directly and very accurately by finding the frequencies which induce transitions between the Zeeman levels. If the ground state of the ion contains a number of levels which are separated in zero magnetic field (within a few cm^{-1}), these separations can also be determined accurately.

Bleaney and Stevens (1953, henceforth referred to as I) have reviewed the detailed theory and the experimental techniques associated with paramagnetic resonance, with particular emphasis on the application of the method to salts containing transition group ions, where the paramagnetism arises from electrons in an unfilled d or f shell. The complementary purpose of the present report is to collect together the available paramagnetic resonance data on these salts, and to present in a fairly simple way such part of the theory as is necessary for an understanding of the results. In general, these data give valuable information on the following properties of transition group ions in crystals:

(i) Under certain conditions the resonance data may be very useful when considering bulk properties of these crystals, e.g. susceptibility and specific heat. This is especially true at very low temperatures, when such properties depend only on the energy levels within a few cm^{-1} of the ground level, i.e. on those levels measured by resonance methods.

(ii) Although the resonance spectrum arises from electronic transitions, information on the nucleus of the paramagnetic ion can be obtained if there is a hyperfine structure arising from interaction between electrons and nucleus. For example, the nuclear spin and approximate values for the nuclear magnetic dipole and electric quadrupole moments may be found.

(iii) The lowest electronic energy levels, and therefore the resonance spectrum, are extremely sensitive to the environment of the paramagnetic ion in the crystal. The spectrum may thus provide information about the symmetry of the surroundings of the ion, and about the nature and strength of the binding between the ion and its immediate diamagnetic neighbours.

(iv) The widths of the absorption lines in a fully resolved spectrum (for more details see I) depend on a number of factors including (*a*) the rate at which the microwave energy absorbed by the electron spins is transferred into thermal (lattice-vibration) energy, i.e. on the spin–lattice relaxation time; (*b*) the magnetic dipole interaction between neighbouring paramagnetic ions; (*c*) the exchange interaction between neighbours. The resonance spectrum may thus give some information on these factors. However, the line widths will not be discussed in this report, mainly because the emphasis in most investigations of transition group ions in crystals has been on obtaining accurate information about the positions of the lowest energy levels. The experiments are then designed as far as possible to minimize all contributions to the line width, in order to obtain maximum resolution of the spectrum.

The arrangement of the report is as follows. First, a general discussion is given in § 2 of the magnetic properties of the transition group salts which have been

measured in resonance experiments. This includes a brief outline of the effects on the magnetic behaviour of crystalline electric fields and covalent bonding. In § 3 the lowest energy levels of the paramagnetic ion, and the associated paramagnetic resonance spectrum, are considered in more detail. In particular, the usefulness of a spin Hamiltonian for describing the energy levels, and hence the spectrum, is emphasized, and a simple explanation of the various terms in this Hamiltonian is given. In § 4 a brief description is given of the crystal structures of the paramagnetic salts most commonly used in resonance experiments, and there is some discussion of the information about crystal structures which can be inferred from the resonance data. Finally, the available resonance data for salts containing ions of the 3d, 4d, 5d, 4f and 5f–(6d) transition groups are tabulated in § 5. These data refer mostly to measurements on magnetically dilute single crystals, because such measurements give the most detailed information about the lowest energy levels.

§ 2. The Paramagnetism of Transition Group Ions in Crystals

In this chapter we shall give a general outline of the paramagnetic properties of the transition group salts which have been measured in paramagnetic resonance experiments. The relevant transition group elements are given in table 2.1; the paramagnetism arises from the unpaired d or f electrons of ions of these elements.

Table 2.1

Name of group	Atomic numbers	Incomplete shell
Iron	22 (Ti) to 29 (Cu)	3d
Palladium	40 (Zr) to 47 (Ag)	4d
Rare earth	58 (Ce) to 70 (Yb)	4f
Platinum	72 (Hf) to 79 (Au)	5d
Actinide	90 (Th) onwards	5f–6d

Fairly typical examples of the crystal types containing such ions which have been used in each group are shown in column 2 of table 2.2. It must be emphasized that for these examples M does not necessarily stand for all of the elements in the corresponding group in table 2.1, although for the ethylsulphates of the 4f group such a complete series of salts can be made.

Table 2.2

Transition group	Example of crystal	Diluting ions	Paramagnetic complex	Crystal field strength	
3d	$K_2M^{II}(SO_4)_2.6H_2O$	Zn^{2+}	$[M(H_2O)_6]^{3+}$	Medium	
	$K_3M^{III}(CN)_6$	Co^{3+}	$[M(CN)_6]^{3-}$	Strong	Covalent
4d	$M^{III}(NH_3)_4Cl_3$	Co^{3+}	$[M(NH_3)_6]^{3+}$	Strong	Covalent
5d	$K_2M^{IV}Cl_6.6H_2O$	Pt^{4+}	$[MCl_6]^{2-}$	Strong	Covalent
4f	$M^{III}(C_2H_5SO_4)_3.9H_2O$	La^{3+}	$[M(H_2O)_9]^{3+}$†	Weak	
5f	$M^{VI}O_2Rb(NO_3)_3$	$(UO_2)^{2+}$	$[(O–M–O)(NO_3)_3]^{-}$	Strong (covalent) from 2O; medium from (NO_3) groups	

† The ethylsulphate groups should probably be included.

In all of these salts, when undiluted the neighbouring paramagnetic ions M are fairly well separated from each other by a distance of the order 6 or 7 Å. There is nevertheless a certain amount of magnetic dipole, and in some cases exchange, interaction between neighbours; these interactions influence the widths of the paramagnetic resonance absorption lines (see I). The dipolar interaction is of order 0·01 cm^{-1} in all the salts; it can be pictured as a random field of approximately 100 gauss at each M ion due to the neighbouring M ions, and the resonance absorption lines have a width of this order of magnitude due to this effect. The exchange interaction between a neighbouring pair of M ions varies considerably from one salt to another. Very broadly speaking, for the examples in table 2.2 it is negligible for the 4f group, $\sim$0·01 cm^{-1} for the 3d group hydrated salts, $\sim$0·1 cm^{-1} for the 3d group cyanides and the 4d group, and larger, perhaps several cm^{-1} in some cases, for the 5d group. It is of interest to note that the increase of exchange interaction follows the increase in the interaction between M and its immediate diamagnetic neighbours (see columns 4 and 5 of table 2.2), which implies that the mechanism of exchange coupling between two M ions involves the diamagnetic ions which lie between them. The point, however, which we wish to emphasize here is that exchange interaction generally leads to an increase in the width of the resonance absorption lines (although for some special cases, e.g. in $CuSO_4.5H_2O$, there may be narrowing: see I). To resolve the resonance spectrum for salts where the exchange is appreciable, as in the 5d group, it is essential to use magnetically dilute mixed crystals, where most of the paramagnetic M ions are replaced by diamagnetic ions. Even when the exchange is small or negligible, as in the 4f group, the width arising from magnetic dipolar interaction often prevents complete resolution of the spectrum, and more information about the energy levels is obtained from measurements on diluted crystals. It should be remarked, however, that the positions of the energy levels in diluted crystals are generally not quite the same as in undiluted ones, owing to slight changes in the distortions of the magnetic complex. Typical examples of suitable diamagnetic ions are given in column 3 of table 2.2; these can be used to dilute the corresponding crystals in column 2.

For diluted crystals, the paramagnetic properties and the resonance spectrum depend only on the particular ion M and the electrostatic interaction between M and its immediate diamagnetic neighbours, i.e. they depend on the paramagnetic complex (see column 4 of table 2.2). The much smaller effect of the more distant neighbours can be neglected (see, for example, Van Vleck 1939). The complexes can be divided into three broad classes according to the strength of the electrostatic interaction, as follows:

(*a*) Weak crystal field, e.g. hydrated 4f group ions. The magnetic 4f electrons lie fairly deep in the ion, and are well shielded from the electric field of the surrounding water dipoles. The effect of the crystal field is then weaker than the Russell–Saunders (*s.s* and *l.l*) and spin–orbit (*l.s*) coupling between the 4f electrons, and the magnetic properties of the ion are in many ways similar to those of a free ion.

(*b*) Medium crystal field, e.g. most hydrated 3d group ions. The magnetic electrons are in 3d orbits, and approach fairly close to the neighbouring water dipoles, so that they are strongly exposed to the electric field of these dipoles. The effect of the crystal field is then considerably greater than the spin–orbit coupling, but less than, or sometimes of the same order as, the Russell–Saunders coupling. This results

in reduction or quenching of the orbital magnetic moment of the 3d electrons, as will be discussed more fully below.

(*c*) Strong crystal field or covalent bonding, e.g. the cyanides of 3d group, and all known complexes of the 4d and 5d groups. For this case the interaction between the magnetic electrons and the attached negative ions is stronger than the Russell–Saunders coupling; this may result in a reduction of the free ion spin magnetic moment, in addition to a reduction in the free ion orbital magnetic moment. For most cases, the reason for the interaction being stronger than in (*b*) is that the d-orbits of the magnetic electrons appreciably overlap the orbits of the electrons on the neighbouring diamagnetic ions.

This distinction between the crystal field strengths (*a*), (*b*) and (*c*) was first given for salts of the 3d and 4f groups by Van Vleck (1932a, 1935) and others, and was based on experimental evidence from susceptibility measurements. Paramagnetic resonance data have confirmed this classification, and have given some detailed information about the crystal field and/or covalent bonding in salts of all five transition groups.

2.1. *The* 3d, 4d *and* 5d *Groups* (*Octahedral Complexes*)

In these groups the experimental data are mostly for complexes where the positively charged paramagnetic ion is at the centre of a nearly regular octahedron of negatively charged diamagnetic ions (or radicals), or of electric dipoles with their negative ends pointing inwards. Examples are $[Cu(H_2O)_6]^{2+}$, which can be written $Cu^{2+}(H_2O)_6$, and where the configuration of magnetic electrons is $Cu^{2+}=3d^9$; also $[IrCl_6]^{2-}$, which can be written $Ir^{4+}Cl^-{}_6$, where $Ir^{4+}=5d^5$. It may be remarked that it is not strictly correct to represent either of these complexes as being purely ionic, although this is quite a good approximation for the copper complex, and for most other hydrated 3d group complexes. However, using at first a purely ionic model, we will now consider the effect of the attached negative ions (or electric dipoles) on the magnetic d-electrons belonging to the central ion. The rigorous theoretical treatment of the crystal field in these complexes is reviewed in I; here we will only outline the theory in fairly simple terms.

Suppose at first that the central ion has just a single electron in a d-shell, with $l=2$, $s=\frac{1}{2}$. There are five d-orbits, corresponding to the $(2l+1)$-fold orbital degeneracy, and the associated wave functions can be expressed as a function of the radius r multiplied by functions with angular variation of the form xy/r^2, yz/r^2, zx/r^2, $(x^2-y^2)/r^2$ and $(2z^2-x^2-y^2)/r^2$ respectively. These five orbitals can be represented diagrammatically by the charge cloud distributions in figure 2.1, where the dots represent the attached negative ions.

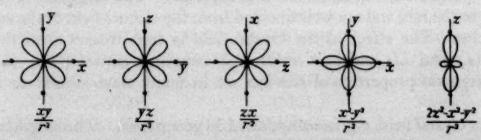


Figure 2.1. Diagrams of the charge clouds associated with the orbitals of a d-electron on a paramagnetic ion. The dots on the axes represent the negative ions (or molecules) of the octahedral complex. When there is covalent bonding, the p (and for some cases s) orbitals of the outer electrons on these negative ions must also be considered.

For a free ion, the energy corresponding to the d-electron being in any one of these five orbits is equal, so that the $^2D_{5/2}$, $^2D_{3/2}$ ground term multiplets contain all of these orbitals equally. On the other hand, when there are negative ions on the $x, y, z, -x, -y, -z$ axes, two things can happen:

(i) The d-electron tries to keep in orbits such that the negative charge cloud avoids the attached negative ions. Hence, the orbitals xy, yz and zx (usually called $d\epsilon$ orbitals) have lower electrostatic energy than the $d\gamma$ orbitals x^2-y^2, $2z^2-x^2-y^2$ (see figure 2.2). This is the origin of the crystal field splitting of the ground term, and leads to quenching of the orbital magnetic moment.

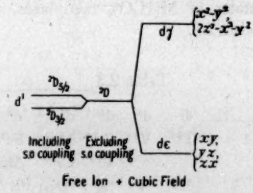


Figure 2.2. The energy levels of a single d-electron on a free ion (e.g. Ti^{3+}, $3d^1$) are on the left. The levels in a cubic (octahedral) crystalline electric field are on the right, where the splittings due to spin–orbit coupling are neglected. The $d\epsilon$ orbitals have lower energy than the $d\gamma$ ones, because the associated charge clouds (figure 2.1) avoid the attached negative charges.

(ii) There may be covalent bonding, which in simple terms can be pictured as meaning that electrons in orbits on the attached negative ions tend to move into orbits on the central positive ion, thereby evening out the charge distribution. This leads to an increased splitting between the $d\epsilon$ triplet and the $d\gamma$ doublet in figure 2.2, and also to other effects which will be mentioned below. (For a detailed discussion of the effects of covalent bonding on the magnetic behaviour, see Van Vleck 1935, Stevens 1953, Owen 1955a.)

Medium crystal field.

In hydrated salts of the 3d group the crystal field is the more important factor, and covalent bonding, although definitely present, only gives a fairly small additional contribution to the splitting. The total splittings are $\sim$10000 cm^{-1} for divalent complexes $[M''(H_2O)_6]^{2+}$, and $\sim$20000 cm^{-1} for trivalent complexes $[M'''(H_2O)_6]^{3+}$. With a single electron on the central ion (d^1), there is then partial quenching of the orbital magnetic moment because two of the orbits ($d\gamma$) are not occupied, and therefore do not contribute to the moment. In fact, the detailed theory (Abragam and Pryce 1951) shows that the orbital magnetic moment associated with the lower triplet ($d\epsilon$) is as if there were an effective orbital quantum number of $l'=1$ rather than $l=2$. For the general case of d^n, the n electrons fill the $d\epsilon$ triplet and the $d\gamma$ doublet in accordance with Pauli's principle, and the lowest energy state is obtained when (*a*) the total spin S is a maximum (Hund's rule) and, when condition (*a*) is fulfilled, (*b*) the minimum number of electrons occupy the high energy $d\gamma$ orbits. (There is a slight exception to (*b*) for d^2 and d^7, but this does not affect the main

argument.) Thus for d^2, the ground state configuration is $(d\epsilon)^2$, the total spin is $S=\frac{2}{2}$, and since the three electrons exactly half fill the $d\epsilon$ triplet, which can be treated like a subshell, the effective total orbital moment is $L'=0$. For d^5, the configuration is $(d\epsilon)^3\,(d\gamma)^2$, $S=\frac{5}{2}$, and since both subshells, $d\epsilon$ and $d\gamma$ (i.e. the complete d-shell), are exactly half full, $L=L'=0$. For d^8, the configuration is $(d\epsilon)^6\,(d\gamma)^2$, so that the $d\epsilon$ subshell is closed and the $d\gamma$ subshell is exactly half full, which again gives $L'=0$, and the total spin $S=1$. The argument outlined above, while not very rigorous, does indicate how for d^3, d^5 and d^8 the cubic crystal field leads to orbital singlet ground states, i.e. zero orbital magnetic moment in the first approximation. Applying similar arguments to the other configurations, it is found that the ground states of $M(H_2O)_6$ complexes, where $M=3d^n$, can be represented as shown in table 2.3.

Table 2.3

Configuration		d^1	d^2	d^3	d^4	d^5	d^6	d^7	d^8	d^9
Examples of ions		Ti^{3+}	V^{3+}	Cr^{3+} V^{2+}	Mn^{3+} Cr^{2+}	Fe^{3+} Mn^{2+}	Fe^{2+}	Co^{2+}	Ni^{2+}	Cu^{2+}
Ground term of free ion		^2D	^3F	^4F	^5D	^6S	^5D	^4F	^3F	^2D
Ground term of ion in complex	total spin S	$\frac{1}{2}$	1	$\frac{3}{2}$	2	$\frac{5}{2}$	2	$\frac{3}{2}$	1	$\frac{1}{2}$
	orbital degeneracy	3	3	1	2	1	3	3	1	2

The effect of the spin–orbit coupling, and any non-cubic part of the crystal field arising from a distortion of the water octahedron, is to produce further splittings of these ground state levels. These splittings are usually of the order of a few 100 cm^{-1} to 1000 cm^{-1} when there is orbital and spin degeneracy, and $\sim$1 cm^{-1} when there is only spin degeneracy, although at least two-fold degeneracy remains if S is half integral (Kramers' theorem). As an illustration, some of these splittings are shown diagrammatically in figure 2.3 for the case where there is a distortion of tetragonal symmetry, i.e. an elongation or compression of the octahedron along one of its axes. Paramagnetic resonance transitions would normally be induced only between the lowest levels in figure 2.3, i.e. those within a few cm^{-1} of the ground level.

Strongly bound complexes.

In the octahedral cyanides of the 3d group, and all known octahedral complexes of the 4d and 5d groups, the magnetic properties can be interpreted by assuming that the separation between the $d\epsilon$ and $d\gamma$ levels is larger than the Russell–Saunders coupling energy. (An approximate measure of this energy is given by the separation between terms of different multiplicity in the free ion.) This means that the reduction in energy which is obtained by keeping all the individual electron spins parallel to give maximum S (Hund's rule) is more than lost if, in order to satisfy Pauli's principle, some of the electrons have to be put in the now very high energy $d\gamma$ orbits. The lowest energy state is then obtained with all electrons of d^n in the $d\epsilon$ subshell, and with $n=6$ this subshell is closed, giving a singlet diamagnetic ground state. For the general case of d^n, the ground states can be found as before, and are shown in table 2.4.

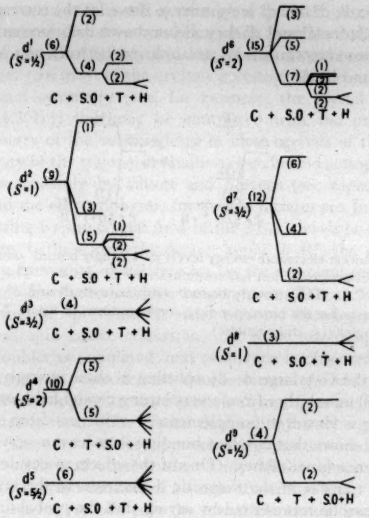


Figure 2.3. Illustrations of the lowest electronic energy levels of paramagnetic ions in tetragonally distorted octahedral complexes in which the cubic crystal field is of 'medium' strength, e.g. most hydrated 3d group complexes. The total degeneracy (spin plus orbit) of each level is shown in parentheses. C=cubic field, which, for each configuration, leaves the ground states given in Table 2.3 ~10000 cm^{-1} below all other levels. S.O=spin–orbit coupling interaction, which produces splittings as shown ~100 cm^{-1}. T=tetragonal part of crystal field, which (*a*) is assumed for simplicity to give splittings small compared with those from S.O for d^1, d^2, d^8 and d^7; (*b*) is given an arbitrary sign—if the sign were reversed, the splitting due to T would be inverted; (*c*) produces small splittings (~1 cm^{-1}) between the lowest levels of d^3, d^4 and d^8 which are not shown in the diagrams. H=applied magnetic field, giving separations between the levels which are greatly exaggerated in the figure. The paramagnetic resonance spectrum arises from transitions between these last levels.

Table 2.4

Configuration		d^1	d^2	d^3	d^4	d^5	d^6
Example of complex				$[Cr(CN)_6]^{3-}$	$[Mn(CN)_6]^{3-}$	$[Fe(CN)_6]^{3-}$	$[Co(CN)_6]^{3-}$
Ground term	total spin S	½	1	3/2	1	½	0
	orbital degeneracy	3	3	1	3	3	1

Strongly bound octahedral complexes with $n>6$ do not appear to be formed, presumably because they would be too unstable with electrons in the high energy $d\gamma$ orbits. As before, the spin–orbit coupling and distortions of the octahedron lift the degeneracy of these ground states. With a tetragonal distortion, the energy

level diagrams for d^1, d^2 and d^3 are similar to those for the corresponding configurations in figure 2.3; for d^4 and d^5 they are as shown diagrammatically in figure 2.4. Paramagnetic resonance transitions are induced between the levels within a few cm^{-1} of the ground level.

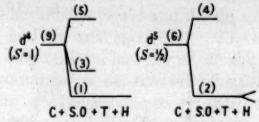


Figure 2.4. The lowest electronic energy levels in 'strongly bound' octahedral complexes in which the paramagnetic ion has configurations d^4 and d^5. The notation is the same as in figure 2.3. For strongly bound complexes d^1, d^2 and d^3, the lowest levels are similar to those for the corresponding configurations in figure 2.3; for d^6 the ground state is a singlet (i.e. diamagnetic).

At first sight the very large $d\epsilon$–$d\gamma$ splitting in these strongly bound complexes might equally well arise either from a very strong crystal field, or from the effect of covalent bonding. However, supplementary evidence from the paramagnetic resonance results shows that covalent bonding is at least a very important factor. Briefly, this evidence is as follows. One of the effects of covalent bonding is that there is a partial transfer of the magnetic d-electrons to the attached ions of the complex. This can be represented by saying that the probability of finding a $d\gamma$ orbit electron on the central M ion is no longer 1 but α^2, ($\alpha^2 \leqslant 1$), and the probability of finding a $d\epsilon$ orbit electron on M is β^2 ($\beta^2 \leqslant 1$). This transfer of the magnetic electrons leads to several effects on the magnetic behaviour, in addition to an increased splitting between the energies of $d\gamma$ and $d\epsilon$ orbitals. Two of the most important effects are: (i) the orbital magnetic moment is slightly smaller than that given by a purely ionic model; (ii) there may be interaction between the magnetic electrons and the magnetic moments of the nuclei of the attached ions; these can then give a hyperfine structure in the paramagnetic resonance spectrum. Experimentally, effect (i) has been observed in a large number of complexes, and effect (ii) in a small number. The results suggest the following orders of magnitude for the amount of charge transfer: 3d group $[M''(H_2O)_6]^{2+}$, $\alpha^2 \sim 0{\cdot}8$, $\beta^2 \sim 1{\cdot}0$; $[M''(H_2O)_6]^{3+}$, $\alpha^2 \sim 0{\cdot}7$, $\beta^2 \sim 0{\cdot}9$; $[M''(CN)_6]^{3-}$, $\alpha^2 \sim 0{\cdot}5$, $\beta^2 \sim 0{\cdot}8$; 4d group $[M''(NH_3)_6]^{3+}$, $\alpha^2 \sim 0{\cdot}5$, $\beta^2 \sim 0{\cdot}9$; 5d group $[M^{iv}Cl_6]^{2-}$, $[M^{iv}Br_6]^{2-}$, $\alpha^2 \sim 0{\cdot}5$, $\beta^2 \sim 0{\cdot}7$. In chemical terminology, $1-\alpha^2$ is a measure of the strength of the σ-bonding, and $1-\beta^2$ is a measure of the strength of the π-bonding. For references and further discussion see Owen (1955a).

2.2. *The* 4f *Group*

For the magnetic complexes of the 4f group which have been investigated in detail by paramagnetic resonance, the effect of the crystal field on the paramagnetic ion is less than in the other groups. This is partly because the average radius of the 4f orbits is small, and partly because the outer closed electron shells screen the 4f shell. The crystal field splitting of the 4f levels is usually of the order 100 cm^{-1}, which is small compared with the spin–orbit interaction (of order 1000 cm^{-1}). J, the

total angular momentum, is thus still quite a good quantum number, and at room temperature the susceptibility does not differ much from that of the free ion.

With one exception (Eu^{2+}) only trivalent 4f ions have been studied by paramagnetic resonance. In most of the crystals investigated, the surroundings of the 4f ion have trigonal symmetry (see, for example, the crystal structure of the ethylsulphates, § 4.3(9)); this may be contrasted with the predominant cubic (octahedral) symmetry of the surroundings in most crystals of the 3d, 4d and 5d groups. The theory of the trigonal crystalline electric field acting on 4f group ions has been developed mainly by Elliott and Stevens (see especially 1953), with special reference to the ethylsulphates; for double nitrates see Judd (1955).

Since the splitting by the crystal field of the $2J+1$ levels of the lowest J-term is of order 100 cm^{-1} (these remarks do not apply to $4f^7$) the susceptibility will depart appreciably from that of the free ion as the temperature is lowered from room temperature. Furthermore, low temperatures (20°K or less) are generally required to examine the spectrum, in order to avoid broadening the absorption lines through rapid spin–lattice relaxation. At these low temperatures, usually only the lowest doublet is populated, and resonance is observed between its two levels. As an illustration, consider the double nitrate of Ce^{3+}, $4f^1$. In the free ion, the two terms belonging to the lowest L–S multiplet are $^2F_{7/2}$ and $^2F_{5/2}$, of which the latter lies lowest. The crystal field splits the six m_J levels of $^2F_{5/2}$ in such a way that a Kramers doublet (approximately characterized by $m_J = \pm\frac{1}{2}$) lies more than 10 cm^{-1} below all other levels. Only this doublet is populated at 4°K, and the resonance spectrum consists of a single absorption line. In a similar way, for most other 4f ions in crystals only the lowest electronic doublet is studied by paramagnetic resonance, and the spectrum consists of a single line, perhaps with hyperfine structure (see § 5.5). An exception is $4f^7$, where the initial splittings between the eight levels belonging to the ^8S ground state are extremely small (of order 0·1 cm^{-1}, or less).

2.3. *The* 5f-(6d) *Group*

Detailed investigations have only been made on crystals containing uranyl-type ions, e.g. $(NpO_2)^{2+}$. In these the paramagnetic atom is strongly bonded to two oxygen atoms, forming a linear O–M–O group; four electrons from the paramagnetic atom are used in the two bonds. In the crystals investigated (double nitrate and double acetate) the surroundings of the O–M–O group have trigonal symmetry about the axis. The detailed theory of Eisenstein and Pryce (1955) shows that the electrons responsible for the paramagnetism are 5f electrons moving in a strong crystal field of trigonal symmetry. For the only examples measured, $5f^1$ and $5f^2$, this field splits the 5f orbital levels in such a way that an electronic doublet lies well below all other levels; the paramagnetic resonance data refer to this doublet.

§ 3. The Lowest Energy Levels and the Paramagnetic Resonance Spectrum

In § 2 we have given a general outline of the magnetic behaviour and the electronic energy levels of the paramagnetic ions in their environment in crystals. Paramagnetic resonance measurements are normally made at temperatures such that only a group of energy levels within a few cm^{-1} of the ground level is populated.

These lowest levels, and the associated resonance absorption spectrum, will now be discussed in more detail. We will first consider a very simple example, and then go on to more complicated cases where the energy levels are most conveniently described by the use of a 'spin Hamiltonian'.

3.1. *The Resonance Phenomenon*

The phenomenon of paramagnetic resonance was first reported by Zavoisky (1945). It is shown in its simplest form by a set of non-interacting paramagnetic ions, each possessing a single unpaired electron and a ' spin only ' magnetic dipole moment of $\frac{1}{2}g_s\beta$, where β is the Bohr Magneton, and $g_s=2{\cdot}00229$. In a steady magnetic field H, each dipole can orient itself parallel or antiparallel to H, with energies $-\frac{1}{2}g_s\beta H$ and $+\frac{1}{2}g_s\beta H$. Magnetic dipole transitions can be induced between these two energy levels by applying a high frequency magnetic field polarized perpendicular to H, and with frequency ν such that the quantum of high frequency energy equals the separation between the levels, i.e. $h\nu=g_s\beta H$. There is then resonance absorption corresponding to the dipoles being flipped from the parallel to the anti-parallel state, and there is induced emission corresponding to the reverse process. If the system remains in thermal equilibrium, more spins are in the parallel state because it is of lower energy, so that there is a net absorption of high frequency power.

In a paramagnetic resonance experiment this absorption can be measured by placing the specimen in a wave guide or cavity resonator, and detecting the drop in the transmitted microwave power at resonance. The normal practice is to use a fixed frequency ν and a variable field H so that, for the example discussed above, if H is steadily increased a single absorption line is found at $H=h\nu/g_s\beta$. With the typical value $\nu=10^4$Mc/s (i.e. wavelength $\lambda=3{\cdot}0$ cm), H is about $3{\cdot}6$ kilogauss. For the resonance measurements to be described below, the wavelengths lie mostly in the range $0{\cdot}5$ cm to 10 cm, values of $1{\cdot}2$ and $3{\cdot}0$ cm being particularly common, and the magnetic field is variable from 0 to about 15 kilogauss. (For a more detailed account of the experimental technique, and for references, see I.)

3.2. *The Spin Hamiltonian*

The spectrum of a transition group ion in a crystal is generally more complex than that for the simple example discussed above. The lowest group of energy levels, i.e. the ones between which microwave transitions are induced, depend in a complicated way on the particular ion and the symmetry and strength of the crystal field (as outlined in § 2), and also on the spin–orbit coupling, and on other effects such as the hyperfine interaction between electrons and nucleus. However, these energy levels, and the experimental resonance data on them, can be described by the use of a ' spin Hamiltonian ' in a fairly simple way which requires no detailed knowledge of the above effects. This will now be discussed.

The lowest group of electronic states is characterized by a single quantum number S', even though each state is really a complicated mixture of the spin and orbital wave functions of the free ion (see Abragam and Pryce 1951). S' is called the effective or fictitious spin of the system, and is defined simply by equating $2S'+1$ to the number of electronic levels in the lowest group. For some cases S' equals the free ion spin S, e.g. for $[Ni(H_2O)_6]^{2+}$, $3d^8$, the lowest orbital level is a

singlet with threefold spin degeneracy (see § 2.1) so that $S=S'=1$. For other cases $S' \neq S$; for example, for many ions with an odd number of electrons the effect of the crystal field and spin–orbit coupling is to leave a Kramers doublet well below all other electronic levels, in which case $S'=\frac{1}{2}$; e.g. for a distorted $[Co(H_2O)_6]^{2+}$ complex, $3d^7$, $S=\frac{3}{2}$ but $S'=\frac{1}{2}$ (see figure 2.3). This use of an effective spin means that for temperatures where only the $2S'+1$ lowest electronic levels are appreciably populated, the paramagnetic ion is treated like a magnetic dipole which has $2S'+1$ allowed orientations in an applied magnetic field, and each energy level is associated with one orientation. But the effective magnetic moment of this dipole (and hence the splittings produced by the applied magnetic field) does not correspond to a 'spin-only' value; it is different by the factor g/g_s, where g is called the spectroscopic splitting factor (often abbreviated to g-value).

The energy levels, however, often have initial splittings in zero magnetic field. This means that the dipole cannot be treated as free but that, in addition to the interaction with the applied magnetic field, energy terms must also be added representing, for example, electrostatic interaction between the magnetic electrons and the crystal field (which may give Stark splittings if $S>\frac{1}{2}$), and magnetic interaction between electrons and nucleus (which may give hyperfine splittings).

The sum of all the terms, written as a sum of energy operators to be applied to the effective spin states, constitutes the spin Hamiltonian $\mathcal{H}$ for the system. The actual energies are then the eigenvalues E_0 which satisfy the operator equation $\mathcal{H}\psi=E_0\psi$, where ψ represents the wave functions of the effective spin states. These energies are usually fairly complicated expressions, and a fuller discussion will be given later.

The advantage of using a spin Hamiltonian is that the rather complicated behaviour of the lowest energy levels of the paramagnetic ion in a magnetic field can be described in a relatively simple way by specifying the effective spin, together with a small number of parameters which measure the magnitudes of the various terms in the Hamiltonian. Each term considered by itself is fairly simple, and will now be discussed in more detail.

(i) *The spectroscopic splitting factor g.*

We will define the g-value in terms of the experimentally observed quantities, i.e. the splittings $g\beta H$ produced between the effective spin levels by an applied magnetic field H. Using a vector model, the magnetic dipole moment associated with the effective† spin S has $2S+1$ allowed orientations with respect to the direction of H, corresponding to magnetic quantum numbers $M=S, S-1, \ldots -S$, and the relative energies are $g\beta HM$. In a resonance experiment the $2S$ allowed transitions ($\Delta M=\pm 1$) between these $2S+1$ equally spaced levels then occur in the same field $H=h\nu/g\beta$, giving $g=21{\cdot}418/H\lambda$, where H is measured in kilogauss and the wavelength λ in centimetres. In practice, these transitions usually do not all occur at the same field, because of the additional energy terms which cause splittings of the levels when $H=0$, and allowance must be made for this when defining g (see below).

This spectroscopic splitting factor is analogous to the Landé splitting factor g_J for a free ion (at least when effective spin equals free ion spin), but in general has

† S rather than S' will be used to denote effective spin, except where there is danger of confusion.

different values for an ion in a crystal; this is because the part of the magnetic moment arising from the orbital motion of the electrons is modified by the crystalline electric field (see § 2). For some cases the orbital moment is greatly reduced by the crystal field, and then g may be fairly close to g_s.

Also, since the magnitude of the orbital part of the magnetic moment depends on the crystal field, its magnitude is usually different for different directions of H, and shows an angular variation which follows the symmetry of the crystal field. The total g-value (spin plus orbit) may then be anisotropic by an amount which depends on the magnitude of the orbital contribution to the moment, and on the asymmetry in the crystal field. For all cases, an anisotropic g-value can be represented by three principal values g_x, g_y, g_z, which are measured in mutually perpendicular directions and include the maximum and minimum values. If H has direction cosines (l, m, n) with respect to (x, y, z), the g value for this direction is

$$g=(l^2g_x^2+m^2g_y^2+n^2g_z^2)^{1/2}.$$

In the spin Hamiltonian, the splitting of the effective spin levels by H is represented by the energy operators $\mathcal{H}=\boldsymbol{\beta}(g_xH_xS_x+g_xH_xS_x+g_yH_yS_y)$. If H is parallel to z for example, the Hamiltonian reduces to $\mathcal{H}=g_z\boldsymbol{\beta}HS_z$, the eigenstates can be represented by the magnetic quantum numbers $M=S, S-1, \ldots, -S$, and the eigenvalues are $g_z\boldsymbol{\beta}HM$. A fuller discussion of the calculation of the energies from the spin Hamiltonian will be given in § 3.3.

(ii) *Fine structure.*

Very often there are initial splittings of the effective spin levels when $H=0$, although there is always at least two-fold degeneracy if the free ion spin is half integral (Kramers theorem, neglecting nuclear interactions). The most important cause of these splittings is the Stark effect arising from asymmetrical crystalline electric fields. For example, a field of axial symmetry may give splittings which can be formally represented by the operator $D\{S_z^2-\frac{1}{3}S(S+1)\}$ in the Hamiltonian, and fields of lower symmetries may give additional splittings represented by $E(S_x^2-S_y^2)$. The term $D\{S_z^2-\frac{1}{3}S(S+1)\}$ implies that, with $H=0$ and $S=1$ for example, the levels characterized by $M=\pm1$ are separated from the level $M=0$ by $D(1^2-0^2)=D$; similarly with $H=0$ and $S=\frac{3}{2}$, the levels $M=\pm\frac{3}{2}$ are separated from the levels $M=\pm\frac{1}{2}$ by $D[(\frac{3}{2})^2-(\frac{1}{2})^2]=2D$. The Hamiltonian

$$\mathcal{H}=g_z\boldsymbol{\beta}HS_z+D\{S_z^2-\tfrac{1}{3}S(S+1)\},$$

with $S=1$ and $H \parallel z$, thus gives energies $\pm g_z\boldsymbol{\beta}H+\frac{1}{3}D$ for the $M=\pm1$ levels, and energy $-\frac{2}{3}D$ for the $M=0$ level (see figure 3.1). The allowed transitions $\Delta M=\pm1$ then occur in fields H such that $g_z\boldsymbol{\beta}H\pm D=h\nu$, so that there is a fine structure in the spectrum, consisting of two absorption lines separated in magnetic field by $2D/g_z\boldsymbol{\beta}$. Experimentally, the Stark splittings of the lowest levels of transition group ions in crystals are found to have values ranging between zero and a few cm^{-1}.

(iii) *Magnetic hyperfine structure.*

If the nucleus of the paramagnetic ion has magnetic moment μ_N, there is magnetic interaction between μ_N and the electronic magnetic moment μ of energy $E_N=\alpha\mu\mu_N\overline{r^{-3}}$, where r is the radius of the electron orbit and α is a factor, with value usually between 0 and 1, which depends on the ground state wave function (see

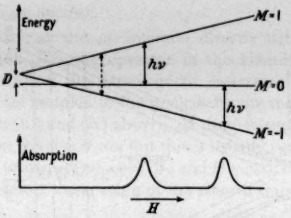


Figure 3.1. The electronic energy levels of the ground state, and corresponding resonance absorption spectrum, for a Ni^{3+}, $3d^8$ ion in a cubic crystal field, with a small component of axial symmetry which splits the spin triplet. The dotted transition is forbidden because it corresponds to $\Delta M=\pm 2$. The figure applies for $H \| z$.

Abragam and Pryce 1951). Putting $\mu_N = Ig_N\beta_N$ (where I=nuclear spin, g_N= nuclear g-value, β_N=nuclear magneton), and $\mu \sim 2\beta S$, this interaction energy can be written $E_N = a2g_N\beta_N\beta\overline{r^{-3}}SI = ASI$. Since the nucleus has $2I+1$ orientations in the magnetic field arising from each of the $2S+1$ orientations of the electronic moment, each electronic energy level is subdivided into $2I+1$ components. The spacings between adjacent components are approximately equal if the applied magnetic field H is strong, but may have a magnitude which is different for different directions of H. In general this anisotropy in A is present if the g-value is anisotropic, but this is not a necessary condition. The hyperfine structure spacing can be described by three principal values A_x, A_y and A_z.

The operator in the Hamiltonian representing this magnetic hyperfine interaction is $A_xS_xI_x + A_xS_xI_x + A_yS_yI_y$. For example, consider the 'isotropic' Hamiltonian $\mathcal{H} = g\beta\mathbf{H}.\mathbf{S} + A\mathbf{S}.\mathbf{I}$, with $S=\frac{1}{2}$, $I=\frac{3}{2}$, $H \| z$ and $g\beta H \gg A$. The $M=+\frac{1}{2}$ electronic level then has energies $\frac{1}{2}g\beta H + \frac{1}{2}Am$, where the nuclear magnetic quantum number m can take values $\frac{3}{2}$, $\frac{1}{2}$, $-\frac{1}{2}$, $-\frac{3}{2}$, and the $M=-\frac{1}{2}$ level has energies $-\frac{1}{2}g\beta H - \frac{1}{2}Am$ (see figure 3.2). The selection rules for transitions are $\Delta M=\pm 1$, $\Delta m=0$, so that the spectrum contains lines at values of H such that $g\beta H + Am = h\nu$, i.e. there are $2I+1=4$ lines with equal spacing A. It will be noted in figure 3.2 that when $H=0$ there are just two degenerate groups of levels, corresponding to the values of $F=I+S=2$ and $I-S=1$, and with degeneracies $2F+1=5$ and 3.

In general, I can always be found from the resonance spectrum simply by counting the $2I+1$ hyperfine lines, and μ_N can be estimated from the measured value of A if $a\overline{r^{-3}}$ (see above) can be calculated. Also, if the nucleus possesses two isotopes with different magnetic moments, the ratio of the moments can be measured accurately.

Experimentally, the structure constant A is found to have values between 10^{-1} and 10^{-3} cm^{-1} for many transition group ions in crystals.

(iv) *Quadrupole interaction.*

The nucleus of the paramagnetic ion may have an electric quadrupole moment Q, and if it is in an electric field gradient $\partial E/\partial z$, some orientations of the nucleus are energetically more favourable than others by an amount of order $eQ\partial E/\partial z$. The greatest contribution to the field gradient (if any) at the nucleus comes mostly from the unpaired magnetic electrons (the contribution from the crystal field is negligible),

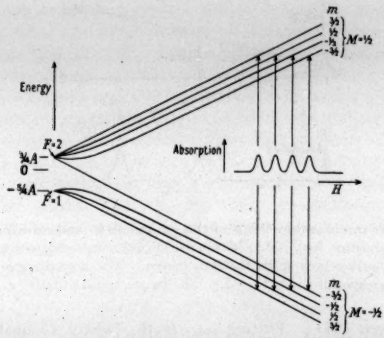


Figure 3.2. The energy levels and the absorption spectrum of an ion with $S=\frac{1}{2}$, $I=\frac{3}{2}$, for which there is isotropic hyperfine splitting and no quadrupole interaction. When $H=0$, or is very small ($g\beta H \ll A$), F is a good quantum number and the levels can be characterized by $F=S+I=2$; and $F=I-S=1$. When $g\beta H \gg A$, as is often the case in paramagnetic resonance experiments, the members of the upper and lower groups of hyperfine levels have equal spacings of $A/2$. The allowed transitions are those for which $\Delta M=\pm 1$, $\Delta m=0$. For this particular example the exact positions of the levels are $\frac{3}{4}A\pm\frac{1}{2}G$, $-\frac{1}{4}A\pm(A^2+\frac{1}{4}G^2)^{1/2}$, $-\frac{1}{4}A\pm\frac{1}{2}(4A^2+2AG+G^2)^{1/2}$, $-\frac{1}{4}A\pm\frac{1}{2}(4A^2-2AG+G^2)^{1/2}$, where $G=g\beta H$ (see p.321).

so that there can be an electrostatic interaction between electrons and nucleus which modifies the energies of the electronic levels. This is represented by operators such as $P\{I_z^2-\frac{1}{3}I(I+1)\}$ in the spin Hamiltonian. For example, consider $\mathcal{H}=g\boldsymbol{\beta}\mathbf{H}.\mathbf{S}+P\{I_z^2-\frac{1}{3}I(I+1)\}$, with $S=\frac{1}{2}$, $I=\frac{3}{2}$ and $H\parallel z$. The energies of the $M=\pm\frac{1}{2}$ electronic levels are then $\pm\frac{1}{2}g\boldsymbol{\beta}H+P(m^2-\frac{5}{4})$, where $m^2=(\pm\frac{3}{2})^2$, $(\pm\frac{1}{2})^2$, so that each electronic level is split into two components with spacing $2P$, corresponding to the different energies associated with different orientations of I in the electric field gradient. It will be noted that the levels $m=\pm\frac{3}{2}$ have the same energy, and so do the levels $m=\pm\frac{1}{2}$. This is simply because the quadrupole moment, which can be represented by two point negative charges with a double positive charge midway between, has reflection symmetry in a plane. If the electric field gradient has less than axial symmetry, an additional term $P'(I_x^2-I_y^2)$ must be added to the Hamiltonian.

The effect of quadrupole interaction on the resonance spectrum is usually complicated because it is accompanied by a larger magnetic hyperfine structure interaction, so that there is then competition between the electric field gradient, and the magnetic field from the unpaired electrons, in fixing the orientations of the nucleus. The result is that, for a general direction of H, the intensities and spacings of the usual $2I+1$ hyperfine lines corresponding to transitions with $\Delta m=0$ are modified by P, and there are also additional lines corresponding to the normally forbidden transitions with $\Delta m=\pm 1$, ± 2. Quadrupole interaction effects have so far been analysed only in hydrated cupric complexes ($P\sim 10^{-3}$ cm^{-1}), in erbium ethylsulphate ($P\sim 3\times 10^{-3}$ cm^{-1}), and in a neptunyl complex ($P\sim 3\times 10^{-2}$ cm^{-1}).

(v) *The direct effect of H on nucleus.*

The nuclear magnetic moment μ_N interacts directly with the applied field H with energy $\sim\mu_N H = g_N\beta_N HI$; the operator in the Hamiltonian is $-g_N\beta_N\mathbf{H}.\mathbf{I}$. This interaction cannot affect the paramagnetic resonance spectrum, unless the paramagnetic electrons are coupled to the nucleus by the magnetic or electrostatic interactions discussed in (iii) and (iv) above. If there is such coupling, there may be a measurable effect on the spectrum, but this is usually very small, or negligible, compared with all other terms; e.g. if $\mu_N = 1\beta_N$ and $H = 3$kG, then $\mu_N H < 10^{-4}$ cm^{-1}. In any case, it will not affect the spacing of the $\Delta m = 0$ transitions, which are the ones normally observed.

The various terms discussed above are in most cases sufficient to describe the lowest energy levels of a paramagnetic ion in a crystal. Collected together, they give the complete spin Hamiltonian:

$$\begin{aligned}\mathcal{H} = {} & \beta(g_z H_z S_z + g_x H_x S_x + g_y H_y S_y) + D\{S_z^2 - \tfrac{1}{3}S(S+1)\} + E(S_x^2 - S_y^2) \\ & + A_z S_z I_z + A_x S_x I_x + A_y S_y I_y + P\{I_z^2 - \tfrac{1}{3}I(I+1)\} + P'(I_x^2 - I_y^2) \\ & - g_N\beta_N \mathbf{H}.\mathbf{I}.\end{aligned}$$

The principal axes (x, y, z) have been taken to be the same for all terms; this is generally found to hold, since their directions are determined by the crystal field. No exceptions have been reported for the crystals considered in this article.

The experimental data on the levels can be represented by appropriate values of the effective electronic spin S' and the nuclear spin I, the spectroscopic splitting factor g, the initial splitting parameters D and E, the magnetic hyperfine splitting parameter A, the quadrupole interaction splitting parameters P and P', and the nuclear g-value g_N. For some ions additional terms are required; these will be described in § 5 under the heading of the particular ion where they are necessary.

The form of the Hamiltonian for particular cases depends largely on the symmetry of the crystal field, or of the surroundings of the paramagnetic ion. The Hamiltonian given above is for the most general case. However, if the symmetry is exactly cubic, as it would be, for example, if the surroundings consisted of a regular octahedron of negative charges, then $g_x = g_y = g_z$, $A_x = A_y = A_z$, $D = E = P = P' = 0$. Another case, which is quite common in practice, is that the symmetry is axial; then $g_z = g_\parallel$, $g_x = g_y = g_\perp$, $A_z = A$, $A_x = A_y = B$, and $E = P' = 0$. This can arise, for example, with a crystal field of trigonal or tetragonal symmetry. In general, the symmetry of the Hamiltonian which describes the resonance data on a particular paramagnetic ion gives a rather sensitive, qualitative, indication of any distortions from a regular arrangement of the surrounding diamagnetic ions; this point will be discussed in § 4.1.

3.3. *Derivation of Energy Levels from Spin Hamiltonian*

We conclude this chapter with a brief explanation of how the energy levels of the ground state are found from a given spin Hamiltonian. We will only outline the rules for making this calculation, without saying anything about the underlying theory. For a more detailed and rigorous account of particular aspects of the theory, and for a derivation of the spin Hamiltonian, reference should be made, for example, to I, to Pryce (1950), Abragam and Pryce (1951), Stevens (1952) and, for a more general theoretical background, to Van Vleck (1932a).

The problem is to find the eigenvalues E_0 which satisfy $\mathcal{H}\psi=E_0\psi$, where ψ represents the wave functions of the effective spin states. The essential steps in the calculations are as follows, where it is at first assumed that $I=0$, i.e. no nuclear interaction terms.

(i) With a strong magnetic field parallel to z, and with z chosen as axis of quantization, the effective electronic spin states can be characterized by the magnetic quantum numbers M. The eigenfunctions are written in the form $|M\rangle$, giving $2S+1$ states $|S\rangle, |S-1\rangle, \ldots, |-S\rangle$. If there are zero field splittings between the levels, then with a weak magnetic field parallel to z, each state is in general of the form $|\psi\rangle=a|S\rangle+b|S-1\rangle+\ldots$, where $a^2+b^2+\ldots=1$. One way of formulating the problem of finding the eigenvalue for a state is that values $a, b, \ldots$ must be found such that the operation $\mathcal{H}|\psi\rangle$ reproduces the same wave function $|\psi\rangle$ multiplied by a number, this number being the required eigenvalue. However, as will be seen below, it is not in fact necessary to know $a, b, \ldots$, in order to find the energy.

(ii) The rules for using the spin operators S_x, S_y and S_z on these states, with z as axis of quantization, are as follows: (*a*) $S_z|M\rangle=M|M\rangle$, i.e. the operation reproduces the same wave function multiplied by a pure number M. (*b*) S_x and S_y are simplest to use when arranged in the form

$$S_+=S_x+iS_y,\ S_-=S_x-iS_y;$$

$$\text{then } S_+|M\rangle=\{S(S+1)-M(M+1)\}^{1/2}|M+1\rangle$$

$$\text{and } S_-|M\rangle=\{S(S+1)-M(M-1)\}^{1/2}|M-1\rangle.$$

When several operations are to be performed in succession, the commutation rules ($S_xS_y-S_yS_x=iS_z$, etc.) must be observed.

(iii) The method for finding the eigenvalues can be shown most easily by an example. Suppose we wish to find the eigenvalues of

$$\mathcal{H}=\beta(g_zH_zS_z+g_xH_xS_x+g_yH_yS_y)+D\{S_z^2-\tfrac{1}{3}S(S+1)\}+E(S_x^2-S_y^2)$$

with $S=1$, when $H\|z$, x and y. First consider $H\|z$, and let the axis of quantization be z. The first term reduces to $g_z\beta HS_z$, and the term in E can be written in the more convenient form $\frac{1}{2}E(S_+^2+S_-^2)$, so that $\mathcal{H}=g_z\beta HS_z+D(S_z^2-\frac{2}{3})+\frac{1}{2}E(S_+^2+S_-^2)$. Operating on each strong field state gives

$$\mathcal{H}_z|1\rangle=(g_z\beta H+\tfrac{1}{3}D)|1\rangle+E|-1\rangle$$

$$\mathcal{H}_z|0\rangle=-\tfrac{2}{3}D|0\rangle$$

$$\mathcal{H}_z|-1\rangle=(-g_z\beta H+\tfrac{1}{3}D)|-1\rangle+E|1\rangle$$

and hence the energy matrix

	$\|1\rangle$	$\|0\rangle$	$\|-1\rangle$
$\|1\rangle$	$g_z\beta H+\frac{1}{3}D$	0	E
$\|0\rangle$	0	$-\frac{2}{3}D$	0
$\|-1\rangle$	E	0	$-g_z\beta H+\frac{1}{3}D$

The eigenvalues are the roots W of the associated secular determinant formed by adding $-W$ to each diagonal element, giving $W_{1,2}=\frac{1}{3}D\pm(g_z^2\beta^2H^2+E^2)^{1/2}$, and $W_3=-\frac{2}{3}D$, for the energies with $H\|z$.

With $H\|x$, the eigenvalues are obtained immediately from those for $H\|z$ by

replacing g_z by g_x, D by $\frac{1}{2}(3E-D)$ and E by $-\frac{1}{2}(D+E)$. This can be seen if, using the relation $S_x^2+S_y^2+S_z^2=S(S+1)$, one rewrites the operator

$$D\{S_z^2-\tfrac{1}{3}S(S+1)\}+E(S_x^2-S_y^2)$$

$$\text{in the form } D_xS_x^2+D_yS_y^2+D_zS_z^2,\ (D_x+D_y+D_z=0)$$

where $D=\frac{3}{2}D_z$ and $E=\frac{1}{2}(D_x-D_y)$. It follows that for $H\|x$, D is replaced by $\frac{3}{2}D_x=\frac{1}{2}(3E-D)$, and E by $\frac{1}{2}(D_y-D_z)=-\frac{1}{2}(D+E)$. Similarly, with $H\|y$, g_z is replaced by g_y, D by $-\frac{1}{2}(D+3E)$ and E by $\frac{1}{2}(D-E)$. These transformation rules are independent of the value of S (those for P, P' are analogous to those for D, E).

(iv) We now consider the case where $I\neq 0$, and there is hyperfine interaction. In a strong magnetic field each effective spin state $|M\rangle$ now consists of $2I+1$ states which are written in the form $|M, m\rangle$, where m is the nuclear magnetic quantum number and can take values $m=I, I-1, \ldots, -I$. The rules for operating with I_x, I_y and I_z are completely analogous to those for S_x, S_y and S_z. If both electron spin and nuclear spin operators are present, each part of the total state function is operated on separately, e.g. $S_zI_z|M, m\rangle=S_zm|M, m\rangle=Mm|M, m\rangle$ and

$$S_+I_-|M, m\rangle=S_+\{I(I+1)-m(m-1)\}^{1/2}|M,m-1\rangle$$
$$=\{S(S+1)-M(M+1)\}^{1/2}\{I(I+1)-m(m-1)\}^{1/2}|M+1, m-1\rangle.$$

The energy matrix is set up in the same way as described above, and is now of order $(2S+1)(2I+1)$. Consider the example

$$\mathcal{H}=g_\parallel\beta H_zS_z+g_\perp\beta(H_xS_x+H_yS_y)+AS_zI_z+B(S_xI_x+S_yI_y),$$

with $S=\frac{1}{2}$ and $I=\frac{3}{2}$, and suppose we wish to find the eigenvalues for $H\|z$. The axis of quantization is chosen to be z and the Hamiltonian can be written

$$\mathcal{H}=g_\parallel\beta HS_z+AS_zI_z+\tfrac{1}{2}B(S_+I_-+S_-I_+).$$

The energy matrix is as follows, where $G=g_\parallel\beta H$, and rows and columns with the same value of $M+m$ are adjacent:

$\lvert M, m\rangle$	$\lvert\frac{1}{2}, \frac{3}{2}\rangle$	$\lvert\frac{1}{2}, \frac{1}{2}\rangle$	$\lvert-\frac{1}{2}, \frac{3}{2}\rangle$	$\lvert\frac{1}{2}, -\frac{1}{2}\rangle$	$\lvert-\frac{1}{2}, \frac{1}{2}\rangle$	$\lvert\frac{1}{2}, -\frac{3}{2}\rangle$	$\lvert-\frac{1}{2}, -\frac{1}{2}\rangle$	$\lvert-\frac{1}{2}, -\frac{3}{2}\rangle$
$\lvert\frac{1}{2}, \frac{3}{2}\rangle$	$\frac{1}{2}G+\frac{3}{4}A$							
$\lvert\frac{1}{2}, \frac{1}{2}\rangle$		$\frac{1}{2}G+\frac{1}{4}A$	$\frac{\sqrt{3}}{2}B$					
$\lvert-\frac{1}{2}, \frac{3}{2}\rangle$		$\frac{\sqrt{3}}{2}B$	$-\frac{1}{2}G-\frac{3}{4}A$					
$\lvert\frac{1}{2}, -\frac{1}{2}\rangle$				$\frac{1}{2}G-\frac{1}{4}A$	B			
$\lvert-\frac{1}{2}, \frac{1}{2}\rangle$				B	$-\frac{1}{2}G-\frac{1}{4}A$			
$\lvert\frac{1}{2}, -\frac{3}{2}\rangle$						$\frac{1}{2}G-\frac{3}{4}A$	$\frac{\sqrt{3}}{2}B$	
$\lvert-\frac{1}{2}, -\frac{1}{2}\rangle$						$\frac{\sqrt{3}}{2}B$	$-\frac{1}{2}G+\frac{1}{4}A$	
$\lvert-\frac{1}{2}, -\frac{3}{2}\rangle$								$-\frac{1}{2}G+\frac{3}{4}A$

Although the energy levels can often only be found by approximate methods, in the present simple example they can be found exactly, since the secular determinant factorizes. It often happens in practice that $g\beta H \gg A,B$, i.e. the

splitting of the electronic levels by the magnetic field is much greater than the hyperfine splitting. In this case the off-diagonal elements W_{ij} can be neglected in a first approximation, since they couple states separated by $\sim g\beta H$, and hence affect the energy only by $\sim(A^2,B^2)/g\beta H$. In a second approximation, the energies are

$$W_{ii}+\Sigma|W_{ij}^2|/(W_{ii}-W_{jj}).$$

§ 4. Crystallography

The detailed interpretation of a paramagnetic resonance spectrum requires the use of a single crystal. The results are closely related to the crystal structure, in particular to the magnetic complex. In this section we shall give some details concerning the structure of more commonly used crystals, and the relation between the structure and the spectrum.

The investigation of single crystals by paramagnetic resonance may be regarded as complementary to their x-ray investigation. Using x-rays, the whole structure of the crystal is determined, and the absolute value of the distance between the atoms measured; paramagnetic resonance cannot contribute to this field of study. On the other hand, if the crystal contains a paramagnetic complex, e.g. a cupric ion surrounded by six water molecules $[Cu(H_2O)_6]^{2+}$, then some properties of the complex can be inferred from the resonance results. In particular, the symmetry of the complex can be inferred, and distortions so small that they have not been observed by x-ray methods are often noticeable. This constitutes the main contribution of paramagnetic resonance to crystallography, although sometimes the results also give some indication of the nature of the atoms or molecules surrounding the paramagnetic ion. More commonly, however, it is the other way round; the x-ray determination of the surroundings aids the interpretation and understanding of the resonance results.

Some crystal structure data will be given for those crystals whose paramagnetic resonance spectra have been reasonably fully analysed. Where a number of members of an isomorphous series have been investigated by the resonance method, the structure of the crystals is briefly discussed in this section. At the same time we will describe the relation of the axes of the paramagnetic complex(es) in unit cell (which are determined by the resonance experiments) with the principal paramagnetic susceptibility axes, and with the crystallographic axes. Some brief crystal structure data for other crystals will be given in § 5, after the tabulation of the appropriate resonance results.

4.1. *The Octahedral Complex*

In the majority of the crystals which have been used in the study of the 3d, 4d and 5d groups, the magnetic complex may be regarded as approximately octahedral, with the central atom (M) surrounded by six neighbours (X). Typical examples of the neighbours are: (1) water molecules, each distant about 2Å from M, (2) halogen ions, each distant about 2·5Å, (3) CN groups, with the carbons distant about 1·8Å and the nitrogens about 3·0Å. It should, however, be remarked that in some cases (in particular Cu^{2+}) the chemical evidence suggests a planar complex of four neighbours instead of an octahedral complex of six, i.e. the effect of two neighbours, which are somewhat more distant from the central ion than the other

four, is neglected in the first approximation. On the other hand, the interpretation of the resonance results, as understood at present, in most cases favours the concept of an octahedral complex. This concept will therefore be used by us and, since it is relevant to so many crystals, the distorted octahedral complex MX_6 will now be considered in some detail, indicating the relationship between the symmetry of the complex and the symmetry of the paramagnetic resonance spectrum.

Let the central ion (M) be at the origin of a Cartesian system of coordinates (x', y', z'), and let the six neighbours (X) be at $(\pm p, 0, 0)$, $(0, \pm q, 0)$, $(0, 0, \pm r)$. Let us consider the various symmetries of the complex.

(*a*) *Cubic.* $p=q=r$, giving a regular octahedron. The resonance spectrum is isotropic, except in two types of special cases, where it has only cubic symmetry. These cases are: (1) $S \geqslant 2$, e.g. Mn^{2+}, Fe^{3+}; (2) hyperfine structure due to X, e.g. $[IrCl_6]^{2-}$, but even in these two cases g is isotropic.

(*b*) *Tetragonal.* $p=q \neq r$, giving an elongation or compression of the octahedron along the z'-axis. The spectrum is axially symmetric in the $x'y'$ plane, except for the two special cases mentioned above (for the special cases, the spectrum has only tetragonal symmetry, but g has axial symmetry).

(*c*) *Trigonal.* The octahedron is stretched or compressed along a line through the centres of opposite faces, e.g. along [111]. The spectrum, just as for (*b*) above, has axial symmetry about the axis of the distortion, except for the special cases.

(*d*) *Orthorhombic.* $p \neq q \neq r$. The spectrum has rhombic axes along x', y', z'.

(*e*) *Lower than orthorhombic.* Such distortions may arise if the atom or molecule at, say, $(+p, 0, 0)$ is not at the same distance from M as that approximately at $(-p, 0, 0)$, or if they lie off the axis, or if they are not of the same nature. The spectrum will still have orthogonal rhombic axes, although not necessarily along x', y', z'.

It is thus seen that paramagnetic resonance, although a powerful means of revealing whether the symmetry of the complex is truly cubic, tetragonal or trigonal, in general does not distinguish between orthorhombic and lower symmetries.

The distortions of the MX_6 complex which occur in practice depend partly on the environment of the complex in the crystal, and partly on the nature of M itself. As an example of the influence of the environment, one may cite the case of the $[Ni(H_2O)_6]^{2+}$ complex in the fluosilicate and Tutton salt, where the distortions have respectively trigonal and orthorhombic, or lower, symmetry. As a striking example of the effect of M itself, one may cite the anomalous results obtained with the $[Cu(H_2O)_6]^{2+}$ complex in crystals with trigonal symmetry (see § 5.3, Cu^{2+}).

4.2. *The Axes*

When discussing the paramagnetism of a single crystal, one must distinguish clearly between three different sets of axes, namely the crystallographic axes, the susceptibility axes, and the axes of the magnetic complexes described in § 3.2.

(1) The crystallographic axes (a, b, c): these are, of course, not necessarily mutually perpendicular. With the help of the diagrams given, for instance, by Groth (1906, 1908, 1910), it is often possible to recognize these axes in a given crystal without detailed goniometric examination.

(2) The principal paramagnetic susceptibility axes (K_1, K_2, K_3): these form a set of orthogonal axes. In suffixes, they are abbreviated (1, 2, 3), e.g. the g-value

along K_1 is written g_1. The susceptibility axes are related to the crystal symmetry, according to the following scheme:

Crystal symmetry	Magnetism	Susceptibility axes
cubic	isotropic	three equal axes arbitrarily oriented
tetragonal, hexagonal, trigonal	axially symmetric about tetragonal, hexagonal, trigonal axis	one axis along tetragonal, hexagonal, trigonal axis; other two equal and arbitrarily situated in perpendicular plane
orthorhombic	—	one along each orthorhombic axis
monoclinic	—	K_2 along diad axis b; K_1, K_3 in ac plane: orientation of those two axes to be found by experiment
triclinic	—	orientation of all three axes to be found by experiment

(3) The principal axes (x, y, z) of each of the magnetic complexes: these are a set of mutually perpendicular axes which may be different for each magnetic complex in unit cell, and whose directions are found from the angular variation of the paramagnetic resonance spectrum. If the unit cell contains a number of magnetic complexes with different orientations, then the axes (x, y, z) of the individual complexes are often not directly related to the morphological axes (a, b, c); the orientation of the axes (x, y, z) must then be determined for each complex, but this is often simplified by axes of some complexes being related to those of others by reflection in a symmetry plane of the crystal (e.g. Tutton salts). The susceptibility axes (K_1, K_2, K_3) are given by the gross effect of the various magnetic complexes in unit cell, and hence are generally related to the axes (x, y, z). If the unit cell contains only one molecule, i.e. only one magnetic complex, then the axes (x, y, z) of the complex coincide with the susceptibility axes of the crystal (e.g. $NiSiF_6.6H_2O$). In a number of crystals, e.g. the ethylsulphates, the unit cell contains more than one molecule, but the magnetic complexes in unit cell are all equivalent and similarly oriented; the axes (x, y, z) then also coincide with the susceptibility axes.

4.3. *Crystal Structures*

Particulars will be given for the following crystals which have been used in detailed paramagnetic resonance investigations:

(1) Tutton salts, $M'_2M''(S^{iv}O_4)_2.6H_2O$, ($S^{iv} \equiv$ S or Se), e.g. $K_2Co(SO_4)_2.6H_2O$.

(2) Alums, $M'M''(S^{iv}O_4)_2.12H_2O$, e.g. $KCr(SO_4)_2.12H_2O$.

(3) Double nitrates, $M''_3M''_2(NO_3)_{12}.24H_2O$, e.g. $Mn_3Gd_2(NO_3)_{12}.24H_2O$.

(4) Fluosilicates, $M''SiF_6.6H_2O$, e.g. $NiSiF_6.6H_2O$.

(5) Bromates, $M''(BrO_3)_2.6H_2O$, e.g. $Cu(BrO_3)_2.6H_2O$.

(6) Sulphates, $M''SO_4.7H_2O$, e.g. $NiSO_4.7H_2O$.

(7) Complex cyanides (*a*) $K_3M''(CN)_6$, e.g. $K_3Fe(CN)_6$.
(*b*) $K_4M''(CN)_6.3H_2O$, e.g. $K_4V(CN)_6.3H_2O$.

(8) Complex halides (*a*) $M'_2M^{iv}X_6$, crystals with the K_2PtCl_6 structure, e.g. K_2IrBr_6.
(*b*) $Na_2M^{iv}X_6.6H_2O$, e.g. $Na_2IrCl_6.6H_2O$.

(9) Ethylsulphates, $M'''(C_2H_5SO_4)_3.9H_2O$, e.g. $Pr(C_2H_5SO_4)_3.9H_2O$.

(10) Uranyl-type rubidium nitrates, $M^{vi}O_2Rb(NO_3)_3$, where $M^{vi} \equiv$ U, Np, Pu, etc.

(M′, M″ and M‴ represent, respectively, a monovalent, divalent and trivalent cation.)

Brief details of some other crystals whose spectra have been studied in detail will be given at the foot of the table in § 5 in which the results for that crystal are recorded.

(1) *Tutton salts.*

The Tutton salts form an extensive isomorphous series with general formula $M'_2M''(S^{iv}O_4)_2.6H_2O$, where M′ is a monovalent cation (e.g. K, Rb, NH_4), M″ is a divalent cation (e.g. Mg, Zn, or any divalent ion of the 3d group), and $S^{iv} \equiv$ S or Se. The crystals are monoclinic (see Tutton 1916). The structure of one member of the series, $(NH_4)_2Mg(SO_4)_2.6H_2O$, has been investigated in detail by Hofmann (1931). There are two molecules in unit cell, and each divalent cation is surrounded by a distorted octahedron of six water molecules; in the crystal investigated four of these lie very nearly in a square, each distant 1·9Å from the ion, and the other two are distant 2·15Å. The dimensions of the octahedron have been criticized by Wyckoff (1951), and in any case one would not expect them to be exactly the same for all members of the series.

The axes (x, y, z) of the two differently oriented, but otherwise equivalent, magnetic complexes $M''(H_2O)_6$ are not related to the monoclinic axes (a, b, c) of the crystal, but the three axes for one complex are derived from those of the other by reflection in the *ac* plane. The experimentally determined orientation of the axes (x, y, z) is often described on the approximation that for each magnetic complex there is an axis of tetragonal symmetry (z). Then α is the angle of inclination of z to the *ac* plane, and the projection of z on the *ac* plane is K_1 or K_2, depending on whether χ_z† is greater or less than $\chi_{x,y}$, since the axes K_1, K_2 are so defined that the susceptibility along K_1 (χ_1) is greater than that along K_2 (χ_2). ψ is the angle between K_1 and c, measured in that sense which makes the angle between a and c obtuse. Assuming axial symmetry $(\chi_\parallel = \chi_z,\ \chi_\perp = \chi_x = \chi_y)$, the principal values of the susceptibility are, taking $\chi_\parallel > \chi_\perp$,

$$\chi_1 = \chi_\parallel \cos^2\alpha + \chi_\perp \sin^2\alpha;\ \chi_2 = \chi_\perp;\ \chi_3 = \chi_\parallel \sin^2\alpha + \chi_\perp \cos^2\alpha;$$

if $\chi_\parallel < \chi_\perp$, the suffixes 1, 2 are interchanged, since $\chi_1 > \chi_2$ by definition of the axes K_1, K_2.

Detailed investigations, however, show that the assumption of an axis of symmetry for the resonance spectrum is not good (see tables, § 5.3), i.e. that the environment of the paramagnetic ion does not have tetragonal symmetry, and that the magnitude, sign and direction of the distortions are often temperature dependent.

(2) *Alums.*

The alums form another extensive isomorphous series; the general formula is $M'M''(S^{iv}O_4)_2.12H_2O$, where M′ is a monovalent cation (e.g. K, NH_4), M″ is a

† χ_z is the gramme ionic susceptibility of one set of ions measured along their z-axis, and not the gramme ionic susceptibility of the whole crystal measured along the z-axis of one of the two sets of ions.

trivalent cation (e.g. Al, or any trivalent ion of the 3d group), and $S^{iv}\equiv S$ or Se. The crystals are cubic, and the susceptibility of a single crystal is therefore necessarily isotropic.

The crystal structures of a large number of alums have been investigated, in particular by Lipson and Beevers (1935) and Lipson (1935). They all have the same space group Pa3, with four molecules in unit cell, but analysis of the complete structure of a number (mainly aluminium sulphate alums) shows that one can distinguish at least three types of structure, known as the α, β and γ structures. The structure is apparently dependent on the radius of the monovalent ion, and one does not expect the substitution of, say, Cr or Fe for Al to change the structure; the crystals which have been investigated by paramagnetic resonance should all belong to the α and β structures.

The detailed structure of a typical alum of the α-type ($KAl(SO_4)_2.12H_2O$) shows that each trivalent ion is surrounded by a nearly regular octahedron of six water molecules. The octahedron has a small trigonal distortion along the crystal [111] axis (the $[1\bar{1}1]$, $[11\bar{1}]$ and $[1\bar{1}\bar{1}]$ axes for the three other molecules in unit cell), and the cubic axes of the octahedron are displaced relative to the crystal cubic axes by rotation through angle ϕ about [111]. The four magnetic complexes in unit cell are equivalent apart from the different orientations of the axes. For the β-alums, however, from the X-ray data it appears that the octahedra are perfectly regular, and that the cubic axes of the octahedra coincide with the crystal cubic axes. Thus for the β-alums the four magnetic complexes in unit cell should be completely equivalent.

The paramagnetic resonance results on α-type ferric alums, diluted with aluminium, are in good agreement with these deductions, but the results on both α- and β-type chromic alums are on the whole not in such good accord. At room temperature, the β-alums seem to have about as great a trigonal distortion as the α-alums, and for one α-alum (undiluted potassium chrome) the spectrum could not be explained in detail by the symmetry assumptions discussed above. At low temperatures, the resonance results indicate crystal transitions in the following undiluted chrome alums; (i) ammonium: sudden transition at about 80°K, below which there are two different types of magnetic complex; (ii) potassium: gradual transition below 160°K, when a second type of magnetic complex appears; (iii) methylamine: transition at about 160°K; at low temperatures the symmetry of the magnetic complex is lower than trigonal.

(3) *Double nitrates.*

The double nitrates form an extensive isomorphous series embracing both the 3d and 4f transition groups. The general formula is $M''_3M'''_2(NO_3)_{12}.24H_2O$, where M″ is a divalent cation (e.g. Mg or any divalent ion of the 3d group), and M‴ is a trivalent cation (e.g. Bi or a trivalent ion of the 4f group). The crystals are trigonal; those investigated generally grow in thick hexagonal plates, with the trigonal axis perpendicular to the plane of the plate.

A preliminary investigation of the crystal structure of cerium magnesium nitrate has been made by Powell (unpublished). The unit cell is rhombohedral, having the form of a cube pulled out along a trigonal axis. There is one trivalent ion in unit cell, situated at the corner of the distorted cube, and divalent ions can be accommodated both at the face centres and at the body centres of the cube.

The resonance results are in accord with this structure; there is only one type of 4f magnetic complex, but both in manganese and in cobalt crystals (each diluted with magnesium) there are two different types of the 3d magnetic complex. These two types have different degrees of anisotropy (i.e. departure from cubic symmetry), and the ratio of the number of more anisotropic ones to less anisotropic appears to be about 1:1·6. (Diluted salts of nickel and copper, however, appear to have only one type.) The resonance results indicate that in both lattice sites the divalent cation is surrounded by a nearly regular octahedron of six water molecules, with a small trigonal distortion whose size is different for the two sites. The resonance results also suggest that the trivalent ion is surrounded principally by nitrate groups.

Owing to the symmetry of the crystal structure, the ' parallel ' and ' perpendicular ' axes of each magnetic complex must coincide with the ' parallel ' and ' perpendicular ' axes of the crystal, and also with the susceptibility axes.

(4) *Fluosilicates.*

The general formula of the fluosilicates is $M''SiF_6.6H_2O$, where M'' is a divalent cation (e.g. Zn or any divalent ion of the 3d group). The crystals are trigonal, and grow in hexagonal pillars parallel to the trigonal axis. The crystal structure of some members of the isomorphous series has been investigated by Pauling (1930); the unit cell is rhombohedral, with one molecule in unit cell. The divalent ion is surrounded by a nearly regular octahedron of six water molecules, with a small trigonal distortion. The ' parallel ' and ' perpendicular ' axes of the magnetic complex coincide with those of the crystal, and also with the susceptibility axes.

The x-ray diffraction data on $MgSiF_6.6H_2O$, however, are incompatible with a unimolecular unit cell (Wyckoff 1951), and the paramagnetic resonance investigation of this crystal, containing some manganese, shows that there are six magnetic complexes in unit cell. Each complex has orthorhombic, or lower, symmetry, and the six complexes are equivalent apart from the different orientations of the distortions.

(5) *Bromates.*

The general formula of the bromates is $M''(BrO_3)_2.6H_2O$, where M'' is a divalent cation (e.g. Zn or any divalent ion of the 3d group). The crystals are cubic, and the susceptibility of a single crystal is therefore necessarily isotropic. The crystal structure of zinc bromate has been investigated by Yu and Beevers (1936); the space group is Pa3, with four molecules in unit cell, and the surroundings of the divalent cation are similar to those of the trivalent ion in the α-alums. Each of the four divalent cations in unit cell is surrounded by a nearly regular octahedron of six water molecules, with a small trigonal distortion about one of the four body diagonals of the unit cube. The resonance results are in accord with this structure.

(6) *Sulphates.*

Crystals of the type $M''SO_4.7H_2O$ belong to two isomorphous groups; the nickel, zinc and magnesium crystals belong to the same one, and are orthorhombic. The structure of the nickel salt has been investigated by Beevers and Schwartz (1935). There are four molecules in unit cell, and each divalent cation is surrounded by a distorted octahedron of six water molecules. The axes (x, y, z) of the four complexes are related to each other by reflection symmetry in the (100), (010), (001) planes.

(7) *Complex cyanides.*

Two different isomorphous series of complex cyanides have been used in the investigation of 3d ions, one for the trivalent ions and the other for the divalent ones ((*a*) and (*b*), respectively, below). From our point of view, the crystals of the two series are rather similar: in each there are only two distinct magnetic complexes in unit cell, each consisting of the paramagnetic ion surrounded by a nearly regular octahedron of six CN groups, and the two magnetic complexes are equivalent apart from the different orientation of the axes. The crystals may be regarded as (pseudo-) orthorhombic and, with at most a very slight approximation, the axes (*x*, *y*, *z*) of one complex are related to those of the other by reflection symmetry in the three mutually perpendicular planes of the crystal. The orientations of the axes (*x*, *y*, *z*) are described in the tables (§ 5.3) by direction cosines defined with respect to the mutually perpendicular crystallographic axes (*a*, *b*, *c*), which are also the principal susceptibility axes.

(*a*) The compounds of the trivalent series which have been used in resonance investigations are $K_3M''(CN)_6$, where M'' is Cr, Mn, Fe or Co. Gottfried and his collaborators (e.g. Gottfried and Nagelschmidt 1930) have shown the structure to be (very nearly) orthorhombic, and have suggested a unit cell of four molecules. Since the morphology is also (pseudo-) orthorhombic, it is for our purposes more convenient to consider this unit cell, regarding the magnetic complexes as equivalent in pairs, rather than the smaller monoclinic unit cell proposed by Barkhatov and Zhdanov (1942), Barkhatov (1942), which contains only two molecules.

(*b*) The compounds of the divalent series which have been used are $K_4M''(CN)_6.3H_2O$, where M'' is V, Mn or Fe. The crystals are monoclinic pseudo-tetragonal, or truly tetragonal. The crystal structure of the monoclinic ferrocyanide (the angle β differs from a right angle by less than $\frac{1}{10}°$), has been investigated by Pospelov and Zhdanov (1947). The unit cell contains four molecules, but the magnetic complexes in unit cell are equivalent in pairs.

(8) *Complex halides.*

In these crystals the paramagnetic atom is surrounded by six halogen atoms, arranged in an octahedron which is sometimes slightly distorted. One must distinguish between two different crystal types, although the magnetic complexes in them are quite similar.

(*a*) $M'_2M^{iv}X_6$, crystals with the K_2PtCl_6 structure. These crystals are cubic, and the susceptibility of a single crystal is therefore necessarily isotropic. The complete structure of $(NH_4)_2IrCl_6$ has been determined by Bokii and Usikov (1940). There are four molecules in unit cell, and each Ir atom is surrounded by a regular octahedron of six chlorine atoms; the four magnetic complexes in unit cell are completely equivalent. The resonance results on $(NH_4)_2IrCl_6$ and K_2IrCl_6, each diluted with the corresponding platinum compound, are in full agreement with this structure. The results on diluted K_2IrBr_6, however, indicate that the octahedron of Br atoms is slightly distorted. For this crystal there appear to be three magnetic complexes in unit cell, and these are equivalent apart from the different orientation of their axes; each octahedron appears to have a small elongation along one of the cubic axes.

(*b*) $Na_2M^{iv}X_6.6H_2O$ (M^{iv}=Ir, Pt; X≡Cl, Br). The crystals are triclinic; the detailed structure has not been determined by x-ray methods, but the resonance

results give considerable information about the magnetic complex. There is only one distinct magnetic complex in unit cell, i.e. if there is more than one molecule in unit cell, then the magnetic complexes in unit cell have identical distortion and orientation. The complex consists of a central atom, say Ir, surrounded by a nearly regular octahedron of six halogen atoms; the octahedron is elongated along the z axis, and so the electric field on the Ir^{iv} atom has predominant cubic symmetry with a small tetragonal component. There is also a slight rhombic distortion, i.e. the Ir–Cl distance is not quite the same for the x and y axes. The orientation of the axes (x, y, z) relative to the triclinic axes (a, b, c) has not been reported.

(9) *Ethylsulphates.*

The general formula of the ethylsulphates is $M'''(C_2H_5SO_4)_3.9H_2O$, where M''' is a trivalent rare earth ion. The crystals are hexagonal, growing in hexagonal prisms parallel to the sixfold axis. The detailed structure has been investigated by Ketelaar (1937). There are two molecules in unit cell, but the two magnetic complexes are completely equivalent, each having trigonal symmetry about an axis parallel to the crystal hexagonal axis. The rare earth ion is at the centre of a triangle of water molecules, and in parallel planes above and below there are two more triangles of water molecules, but rotated through $\pi/3$ with respect to the middle one; the ethylsulphate radicals lie in the middle plane, between the water molecules. (The trigonal symmetry of the crystalline electric field may be contrasted with the predominant cubic symmetry which exists for the ions of the 3d, 5d groups in the crystals discussed above.) The ' parallel ' and ' perpendicular ' axes of the magnetic complex coincide with those of the crystal, and with the susceptibility axes.

(10) *Uranyl-type rubidium nitrate.*

This isomorphous series of hexagonal crystals has been used to study the 5f group. The general formula is $M^{vi}O_2Rb(NO_3)_3$, where $M^{vi} \equiv U$, Np, Pu, etc. The structure has been investigated by Hoard and Stroupe (quoted by Dieke and Duncan 1949). The rhombohedral unit cell contains two molecules, but the two magnetic complexes are completely equivalent. Each complex consists of a linear O–M–O group parallel to the hexagonal axis, and surrounded by three nitrate groups in the equatorial plane. The crystal hexagonal axis is thus a trigonal axis for each complex, i.e. the ' parallel ' and ' perpendicular ' axes of the complex coincide with those of the crystal, and with the susceptibility axes.

§ 5. Experimental Data

This section is divided into six parts. First (§ 5.1) there are some general remarks concerning the formulae and experimental data which are given for each ion, followed by (§ 5.2) a key to the use of this information. The detailed information on the ions is given in four sub-sections: (§ 5.3) the 3d group, (§ 5.4) the 4d and 5d groups, (§ 5.5) the 4f group, and (§ 5.6) the 5f–(6d) group. Each of these sub-sections is prefaced with a short general introduction, and then each ion is considered in turn, giving:

(*a*) the appropriate spin Hamiltonian;

(*b*) where applicable, formulae in terms of the parameters in the spin Hamiltonian for (i) the energy levels in strong field and zero field, (ii) the temperature dependent part of the susceptibility, (iii) the specific heat tail CT^2/R;

(*c*) a table of the experimentally determined parameters for those crystals which have been reasonably fully analysed;

(*d*) a list of other compounds which have been investigated;

(*e*) some data concerning the structure of crystals which have been used in detailed investigations, unless these have already been discussed in § 4.3.

5.1. *The Spin Hamiltonian*

(*a*) For each ion we shall confine ourselves to the simplest spin Hamiltonian (see § 3) which is able to describe reasonably well the available experimental data. It is likely that in a number of cases this Hamiltonian is insufficient. For example, a quadrupole interaction term may be required for cobalt compounds, and the A-value for manganous compounds is probably not accurately isotropic. Such emendations, however, normally would not affect the results seriously; attention will be drawn to the few exceptional cases.

(*b* i) *The energy levels.*

If possible, formulae are given for the energy levels in an applied field of any size, but this can be done only rarely. When this is not possible, formulae are given for the levels in strong field and, as far as possible, in zero field. (Some general formulae, assuming axial symmetry, are given by Bleaney 1951.) However, in some cases where the nuclear interactions are important one cannot give simple formulae even for the zero field levels. Approximate formulae are then given, but in a few cases the approximations are rather crude; to obtain the levels exactly, one must solve the secular equation numerically, often a very laborious procedure.

(*b* ii) *The susceptibility.*

For a magnetically dilute salt where the ground state of the paramagnetic ion is $(2S'+1)$-fold degenerate (S'=effective spin), with $H \| z$ the temperature dependent part of the gramme ionic susceptibility χ_z is $N\beta^2 g_z^2 S'(S'+1)/3kT$, provided that the temperature is such that only this ground state is appreciably populated. For a number of ions, there are often other levels within a few hundred degrees of the ground state, thereby usually radically altering the susceptibility at temperatures of this order. Susceptibility formulae will only be given when such low-lying levels are not expected.

If the ground state is split in zero field, then the susceptibility departs from Curie's law. At temperatures high compared with this splitting, the susceptibility may be evaluated in ascending powers of $1/T$; the term in $1/T$ is as above, and the term in $1/T^2$ will be given in all susceptibility formulae. (A general formula for this term, assuming axial symmetry, is given by Bleaney 1950a.) It may be noted that the hyperfine structure (i.e. A, P) makes no contribution to this term. The term in $1/T^2$ averages to zero for a powder, but introduces anisotropy into the susceptibility of a single crystal, unless the crystal has cubic symmetry. The anisotropy due to this cause is very small at high temperatures, and may be easily outweighed by a small (perhaps undetected) anisotropy in g.

Since paramagnetic resonance gives no information on the diamagnetic correction, and no direct information on the temperature independent paramagnetic term, in all formulae these two contributions (which are often small) will be omitted.

(*b* iii) *The specific heat tail.*

Another consequence of any splitting of the ground state in zero field is a specific heat anomaly. The maximum of this anomaly occurs at temperatures of the order of the splitting, i.e. usually well below 1°K. The high temperature tail of this anomaly can be detected by, for example, paramagnetic relaxation measurements. The tail may be evaluated, in terms of the splitting parameters, in ascending powers of $1/T$; the first and most important term is in $1/T^2$, and formulae for this term will be given. (A formula for the case of axial symmetry is given by Bleaney 1950a.) Following convention, the value of CT^2/R is given, where C is the contribution to the gramme ionic specific heat from the splittings.

(*c*) *The tabulated data.*

If the resonance spectrum has been analysed in some detail, as in most single crystal measurements, the results are fully tabulated; otherwise the compound is only briefly listed (see below), unless the results refer to an ion for which only few resonance data are available. Usually the following data are given in the tables for each compound:

(i) degree of dilution with diamagnetic compound;

(ii) cross reference to the crystal structure;

(iii) temperature of experiment;

(iv) experimentally determined parameters in the spin Hamiltonian;

(v) special features of the results as (*a*) determination of axes of magnetic complex, (*b*) change of spectrum with temperature, (*c*) ratio of the hyperfine structure splittings for two isotopes;

(vi) references.

(*d*) *The listed data.*

Most powder measurements, and also those single crystal measurements which have not been analysed in any detail, are only briefly listed, giving formula, temperature of experiment, (apparent) g-value, and reference. It may be remarked that powder measurements must be treated with considerable caution, since they are liable to a number of errors. To give just one example, if the energy levels of the ground state are already split in zero field, then the deduction of the g-value may be incorrect.

(*e*) *The crystal structures.*

If the resonance spectrum of a single crystal has been fully analysed, the results may be related to the crystal structure (see § 4). Brief details concerning the structure, in particular the magnetic complex, are accordingly given whenever the resonance results are reasonably complete, unless the crystal has already been discussed in § 4.3.

5.2. *Key to Formulae and Data*

Transformation rules.

Formulae for energy levels in an applied field H, and for susceptibility, are given only with $H \| z$. To obtain the corresponding formulae for $H \| x, y$, besides the appropriate change in the suffixes of g and A, the following replacements are made: $H \| x$, D by $-\frac{1}{2}(D-3E)$, E by $-\frac{1}{2}(E+D)$, P by $-\frac{1}{2}(P-3P')$, P' by

$-\frac{1}{2}(P'+P)$; $H\|y$, D by $-\frac{1}{2}(D+3E)$, E by $-\frac{1}{2}(E-D)$, P by $-\frac{1}{2}(P+3P')$, P' by $-\frac{1}{2}(P'-P)$. (The A and B used when there is axial symmetry are $A_\parallel$ and $A_\perp$ respectively, where '$\parallel$' is z.)

Sign convention.

The sign of g is taken to be positive‡, and the other parameters may have either sign. Where these signs have been determined, they are given explicitly in the tables, a '+' sign being prefixed to the numerical value if it is positive; the absence of a sign indicates that it is unknown, and not that it is necessarily positive. Brackets around a sign indicate that it has not been directly determined.

Where a formula contains in a number of places alternative signs, one must take throughout either only the upper signs or only the lower ones, i.e. $\pm a\ \pm b$ stands for $a+b$, $-a-b$, but not for $a-b$, $-a+b$.

Units.

Susceptibility is given in units of e.m.u. per gramme ion; the value of $N\beta^2/3k$ has been taken as $\frac{1}{8}$, an approximation of less than 0·1%.

The units of the tabulated splitting parameters, apart from the dimensionless g, are cm^{-1}. However, for the formulae for the $1/T^2$ term in the susceptibility and for CT^2/R, the parameters must be converted into degrees Kelvin; to do this, the tabulated values in cm^{-1} are multiplied by $hc/k=1{\cdot}438$.

To convert cm^{-1} to kilogauss, the relation $(h\nu=g\beta H)$ is 1 $cm^{-1}\equiv 21{\cdot}4178/g$ kG.

Anisotropy of susceptibility.

To evaluate from the given formulae the anisotropic susceptibility of a single crystal, note must be taken of the crystal structure since, if there is more than one magnetic complex in unit cell, the anisotropy of a single crystal is generally less than that of each individual complex.

References.

In the tables, cross references are given to the crystal structure (C.S.); a number refers to the appropriate part of § 4.3; a small letter refers to a note at the foot of the table.

References for the resonance data will usually be given for the most detailed and the last publications on each crystal. The tabulated references are abbreviated, giving the first two letters of the name of the first mentioned author and, if necessary, a serial number, followed by the year of publication with the initial 19 omitted and, where required, a small letter to distinguish publications of the same year. The letter u signifies unpublished data. Thus, for example, one of the publications by Bleaney and Ingram (1951) is referred to as Bl7 51a, and unpublished work by Trenam is referred to as Tr u.

Abbreviations and symbols.

The following abbreviations and symbols are used in the tables and footnotes: C.S. crystal structure (reference).

‡ The g-value used here is a parameter which describes the experimentally observed quantities (cf. § 3.2). From a strictly theoretical point of view, this should be written $|g|$, since the theoretical g-value may have either sign; only $|g|$ is measured in a paramagnetic resonance experiment.

d. dilution, expressed as the ratio of number of ions of diluent to paramagnetic, e.g. ratio of number of Zn ions to Mn ions; in the chemical formulae, the diluent ion is in italics.

d.c. direction cosines of the axes (x, y, z) of the magnetic complexes, with respect to mutually perpendicular crystallographic axes (a, b, c).

M number of molecules in unit cell, as determined by x-ray methods.

M_m number of distinct magnetic complexes in unit cell, as determined by paramagnetic resonance.

p powder.

T temperature of investigation in degrees Kelvin.

u unpublished.

χ_z temperature dependent part of the gramme ionic susceptibility of one set of ions along their z-axis, in e.m.u.; it is assumed that $kT \gg$ any splitting of the ground state in zero field.

A warning.

Although in a number of cases the parameters for diluted crystals have been determined with high precision, it is often dangerous to use these precise quantities in calculating properties of the undiluted crystals, since the values of some of the parameters have been observed to be dependent on the dilution. It should also be noted that some values are temperature dependent, but the general indications are that they change little (if at all) below 20°K.

5.3. *The* 3d *group*

A general discussion of the magnetism of 3d group octahedral complexes has already been given in § 2.1. Nearly all the detailed results refer to distorted octahedral complexes, and the data for different crystals containing the same complex have much in common. The occasional dissimilar result almost certainly indicates a completely different complex.

Table 5.1 lists the ions of the 3d group which have been investigated; the value of the free ion spin S, the effective spin S', and an order of magnitude for g are

Table 5.1

Number of 3d electrons	Ions	S	S'	g	r.m.s. g
1	Ti^{3+}	$\frac{1}{2}$	$\frac{1}{2}$	1·1—1·3	1·2
	VO^{2+}, Mn^{6+}		$\frac{1}{2}$	2·0[a]	2·0
3	V^{2+}, Cr^{3+}	$\frac{3}{2}$	$\frac{3}{2}$	2·0	2·0
4	Cr^{2+}	2	2	$1{\cdot}9_5$—2·0	2·0
5	Mn^{2+}, Fe^{3+}	$\frac{5}{2}$	$\frac{5}{2}$	2·0	2·0
			$\frac{1}{2}$[b]	0·7—2·6	1·9
6	Fe^{2+}	2	$\frac{1}{2}$[c]	0—9	5
7	Co^{2+}	$\frac{3}{2}$	$\frac{3}{2}$[d]	2·3	2·3
			$\frac{1}{2}$	1·4—7	4·5
8	Ni^{2+}	1	1	2·2—2·3	$2{\cdot}2_5$
9	Cu^{2+}	$\frac{1}{2}$	$\frac{1}{2}$	2·0—2·5	2·2

a complexes almost certainly not octahedral

b $M(CN)_6$ complex

c FeF_6 complex

d tetrahedral $CoCl_4$ complex

All other data refer to hydrated octahedral complexes $M(H_2O)_6$, but the results for some other complexes are quite similar (e.g. $M(CN)_6$ complex of $3d^3$).

also given for each ion. (S' is defined by equating $2S'+1$ to the degeneracy of the ground state, neglecting nuclear interactions.) The range of g-values which is given refers to the extreme values which have been observed for anisotropic complexes. The last column gives the average for the different compounds of the root mean square value of g.

Table 5.2 lists those naturally abundant isotopes of the atoms given in table 5.1 which have a non-zero value for the nuclear spin I. The corresponding hyperfine structure in the paramagnetic resonance spectrum has been observed for all these isotopes, with the exception of the two titanium isotopes, iron and nickel.

Table 5.2

Z	Atom	A	Natural abund. (%)	I
22	Ti	47	7·75	$\frac{5}{2}$
		49	5·51	$\frac{7}{2}$
23	V	50	0·24	6
		51	99·76	$\frac{7}{2}$
24	Cr	53	9·5$_5$	$\frac{3}{2}$
25	Mn	55	100	$\frac{5}{2}$
26	Fe	57	2·17	—
27	Co	59	100	$\frac{7}{2}$
28	Ni	61	1·25	—
29	Cu	63	69·1	$\frac{3}{2}$
		65	30·9	$\frac{3}{2}$

The natural abundances have been taken from the report by Bainbridge and Nier (1954).

$3d^1$ *Titanium* Ti^{3+}, *Vanadyl* VO^{2+}, *Manganese* Mn^{6+}.

A cubic (octahedral) field leaves an orbital triplet lowest, and this is split into three Kramers doublets by fields of lower symmetry and spin–orbit coupling (cf. figure 2.3). For octahedral complexes which are only slightly distorted, the separation between the doublets is small (of order 100 cm^{-1}) and, since at high temperatures rapid spin–lattice relaxation broadens the lines, paramagnetic resonance can only be observed at very low temperatures.

(*a*) Ti^{3+}. The results on the two crystals which have been studied (see table below) are interpreted in terms of the following spin Hamiltonian, with $S=\frac{1}{2}$:

$$\mathcal{H}=g_{\parallel}H\beta_z S_z+g_{\perp}\beta(H_xS_x+H_yS_y).$$

The nuclear interactions have been omitted, since the hyperfine structure has not been observed. The energy levels are simply given by $\pm\frac{1}{2}g\beta H$, where $g^2=g_{\parallel}^2\cos^2\theta+g_{\perp}^2\sin^2\theta$, and θ is the angle which H makes with the axis. For the oxalate, the complex is not octahedral, and low-lying levels are not expected. The susceptibility with $H\|z$ is thus $\frac{3}{32}g_{\parallel}^2/T$.

Ti^{3+}

Formula	T	$g_{\parallel}$	$g_{\perp}$	Remarks	Ref.
$CsTi(SO_4)_2.12H_2O$ C.S.2	4·2 to 2·5	1·25 ±0·02	1·14 ±0·02		Bl2 55
$KTi(C_2O_4)_2.2H_2O$	90, 20	1·86	1·96	$M_m=2$	Bo2 u

(*b*) VO^{2+}, Mn^{6+}. Measurements have been made on some vanadyl compounds; the complexes are almost certainly not octahedral. The g value of MnO_3 has also been measured.

Formula	VO^{2+} T	g	Ref.
$K_3V_2O_3(C_2O_4)_3.4H_2O$ single crystal	290, 90	$g_1=1{\cdot}95_4$ $g_2=1{\cdot}98_5$ $g_3=1{\cdot}96_7$	† Gr u
$VOCl_2$	290	2·00	La 51
$VO_3(C_6H_4)_3N_3(CH)_2$	290	2·02	La 51
$VOSO_4.2H_2O$	290	1·96	Hu3 53
$VOSO_4.5H_2O$	300	$1{\cdot}990_1\pm0{\cdot}002$	Pa u
	70	$1{\cdot}998_7\pm0{\cdot}002$	
	4	$1{\cdot}994\ \pm0{\cdot}004$	
	Mn^{6+}		
MnO_3	290	1·96	La 51

† In a powder diluted with the titanium salt, the hyperfine structure due to ^{51}V ($I=\frac{7}{2}$) was observed at 290°K, 90°K: $A=0{\cdot}010\pm0{\cdot}001$.

$3d^3$ *Vanadium* V^{2+}, *Chromium* Cr^{3+}.

A cubic (octahedral) field leaves an orbital singlet lowest, with fourfold spin degeneracy; this spin quadruplet is split into two Kramers doublets by fields of lower symmetry. The results (see tables below) are interpreted in terms of the following spin Hamiltonian, with $S=\frac{3}{2}$ and appropriate value of I (see table 5.2);

$$\mathcal{H}=g\beta(H_xS_x+H_xS_x+H_yS_y)+D(S_z^2-\tfrac{5}{4})+E(S_x^2-S_y^2)+A(S_xI_x+S_xI_x+S_yI_y).$$

If the last term is neglected, then when $H=0$ there are two degenerate doublets separated by $2(D^2+3E^2)^{1/2}$, and on applying $H\|z$, the levels are at

$$\tfrac{1}{2}g\beta H\pm(\{D+g\beta H\}^2+3E^2)^{1/2},\ -\tfrac{1}{2}g\beta H\pm(\{D-g\beta H\}^2+3E^2)^{1/2}.$$

With $H\|z$, the susceptibility χ_z is

$$\frac{15}{32}\frac{g^2}{T}\left(1-\frac{4D}{5T}\right)\quad (D \text{ in } ^\circ\text{K}).$$

The parameters D and E are sensitive to temperature and dilution.

(*a*) V^{2+}. Owing to the low natural abundance of ^{50}V, we shall consider only ^{51}V ($I=\frac{7}{2}$). With $H\|z$, and $g\beta H\gg E$, A, the levels are at

$$\begin{aligned}&\pm\tfrac{3}{2}g\beta H+D\pm\tfrac{3}{2}Am\\&\pm\tfrac{1}{2}g\beta H-D\pm\tfrac{1}{2}Am\end{aligned}\quad (m=\tfrac{7}{2},\ \tfrac{5}{2},\ \ldots,\ -\tfrac{7}{2}).$$

To apply the transformation rules for $H\|x$, y, the further condition $g\beta H\gg D$ is required.

The expressions for the zero field levels are complicated, the effect of the term in A being to give a spread of levels where previously there was only one. To give an indication of the extent of this spread, let us take $D\gg E$, A. Then if D is

positive, the upper levels are spread over $21A/2$ and the lower over $8A$; this is reversed if D is negative.

The value of CT^2/R is

$$D^2+3E^2+\tfrac{315}{16}A^2 \quad (D, E, A \text{ in } ^\circ\text{K}).$$

(*b*) Cr^{3+}. For 90·4% of the ions there is no hyperfine structure; the values of A for ^{53}Cr ($I=\frac{3}{2}$) are small, and as a result one can often neglect the nuclear interactions.

For ^{53}Cr, the formulae for the strong field levels are the same as for ^{51}V, except that m now only takes the values $+\frac{3}{2}$ to $-\frac{3}{2}$. The zero field levels, assuming $D \gg E$, A, are two singlets at $-D+\frac{7}{4}A$, $-D-\frac{9}{4}A$, and seven doublets at $D\pm\frac{3}{2}Am$, ($m=\frac{3}{2}$, $\frac{1}{2}$, $-\frac{1}{2}$, $-\frac{3}{2}$), $-D+\frac{3}{4}A$, $-D+\frac{1}{2}(\pm\sqrt{13}-\frac{1}{2})A$.

The value of CT^2/R is $D^2+3E^2+0{\cdot}0955(75/16)A$ (D, E, A in °K) where the factor 0·0955 has been introduced to take account of the natural abundance of ^{53}Cr.

$3d^4$ *Chromium* Cr^{2+}.

A cubic (octahedral) field leaves a ' nonmagnetic ' (Van Vleck 1932b) orbital doublet lowest. This doublet is split by a field of tetragonal symmetry (cf. figure 2.3), and paramagnetic resonance is observed for the lower level. (Since the doublet is not split by a trigonal field, it would be interesting to examine the spectrum in a trigonal crystal; cf. Cu^{2+} below.) The fivefold spin degeneracy of the lowest orbital level is completely lifted by a field of rhombic symmetry.

Only one compound, the sulphate, has been investigated in detail. The hyperfine structure of the odd isotope has not been observed, and so the nuclear interactions are omitted below. The results are interpreted in terms of the following spin Hamiltonian, with $S=2$:

$$\mathcal{H}=\beta(g_zH_zS_z+g_xH_xS_x+g_yH_yS_y)+D(S_z^2-2)+E(S_x^2-S_y^2)$$

When $H\|z$, to a good approximation the levels are at $-2D$, $-D\pm(g_z^2\beta^2H^2+9E^2)^{1/2}$, $2D\pm2g_z\beta H$. Owing to the large value of D, the corresponding formula for $H\|x$, y is not a good approximation, since it would require $H\gg20$ kG.

There are five singlets in zero field, at $2D$, $-D\pm3E$, $\pm2(D^2+3E^2)^{1/2}$. With $H\|z$, the susceptibility χ_z is:

$$\frac{3}{4}\frac{g_z^2}{T}\left(1-\frac{7}{5}\frac{D}{T}\right) \quad (D \text{ in } ^\circ\text{K}).$$

The value of CT^2/R is:

$$14(D^2+3E^2)/5 \quad (D, E \text{ in } ^\circ\text{K}).$$

Cr^{3+}

	T	$g_\parallel$	$g_\perp$	D	E	M_m	Remarks	Ref.
$CrSO_4.5H_2O$	290	1·95	1·99	2·24	0·10	2	angle betw. z axes 86°	On2 54

$3d^5$ *Manganese* Mn^{2+}, *Iron* Fe^{3+}.

The lowest level of the free ion is an orbital singlet with sixfold spin degeneracy. For the hydrated complexes, the electric field splits this sextet, giving three Kramers doublets separated normally by less than 1 cm^{-1}. In the cyanides, however, the

V^{2+}

Formula	T	g	D	E	^{50}A	^{51}A	Remarks	Ref.
$(NH_4)_2Zn(SO_4)_2.6H_2O$ C.S. 1; d. 3	290		0·155 ±0·005				$\psi=+4°$; $\alpha=23\frac{1}{2}°$	Hu3 53
d. 1000	20	1·951 ±0·002	0·158 ±0·010	0·049 ±0·005		0·0088 ±0·0002	$\psi=+2°$; $\alpha=22°$ $^{50}A/^{51}A=0·3792\pm0·0008$	Bl8 51 Ki1 53
$K_4[Fe(CN)_6].3H_2O$ C.S. 7; d. 100	90 20	1·992	−0·0264	0·0072	0·0021	−0·0056	$^{50}A/^{51}A=0·380\pm0·001$	Ba4 52

d.c. of axes:		a	b	c
	z	0·470	±0·737	−0·470
	x	0·707	0	0·707
	y	0·523	±0·676	−0·523

Paramagnetic resonance has also been observed in VSO_4, $T=290$; Ba1 48; $V(C_6H_4)_4(CN)_8$, $T=270$ to 20, $g=2·0$, In2 54.

Cr^{3+*}

Formula	T	g	D	E	A	Remarks	Ref.
$CsCr(SO_4)_2.12H_2O$ C.S. 2	290	1·98	$0·072_5$ ±0·003				Ba2 50 a Bl1 50b Ki2 51
	193	1·98 ±0·02	−0·067				
	90, 20	1·98 ±0·02	$-0·066_5$ ±0·001				

Cr^{3+}* (*contd.*)

Formula	T	g	D	E	A	Remarks	Ref.
$KCr(SO_4)_2.12H_2O$ C.S. 2	290	1·98	0·060 ±0·003			Two different magnetic complexes below 160°K; the crystal transition is gradual	Ba2 50a Bl1 50b Ha1 48 Ki2 51 Ti1 51 Wh 48
	193	1·98 ±0·02	0·027 ±0·003				
	160	1·98 ±0·02	$0·017_5$				
	90	1·98 ±0·02	I 0·130 II 0·075 ±0·005				
	20	1·98 ±0·02	I 0·135 ±0·002 II 0·075 ±0·005				
$KAl(SO_4)_2.12H_2O$ C.S. 2; d. 8·5	290	1·98	$0·045_5$				Wh 48
$(NH_3CH_3)Cr(SO_4)_2.12H_2O$ C.S. 2	290	1·98	$0·082_5$ ±0·003			Crystal transition at (157±3)°K; below this temperature, spectrum has rhombic symmetry	Ba2 50a Bl1 50b Ba3 u
	90, 20	$1·97_5$ ±0·01	0·087 ±0·002	0·009 ±0·001			
$(NH_3CH_3)Al(SO_4)_2.12H_2O$ C.S. 2; d. 100	90	$1·97_5$ ±0·01	$0·095_9$ ±0·002	0·009 ±0·001		Crystal transition at (170±3)°K; below this temperature, spectrum has rhombic symmetry	Ba3 u
$(NH_4)Cr(SO_4)_2.12H_2O$ C.S. 2	290	1·98	$0·067_5$ ±0·003			Sudden crystal transition just above 80°K; below this temperature there are two different magnetic complexes	Ba2 50a Bl1 50b Bl11 48 Ha1 48 Ki2 51 Wh 48 Ya 49
	193	1·98 ±0·02	$0·042_5$				
	90	1·98 ±0·02	$0·017_5$				

$(NH_4)Cr(SO_4)_2.12H_2O$ C.S. 2 (*contd.*)	80	1·98 ±0·02	I 0·157 ±0·002 II 0·121 ±0·002	
	20	1·98 ±0·02	I 0·158 ±0·002 II 0·120 ±0·002	
$(NH_4)Al(SO_4)_2.12H_2O$ C.S. 2; d. 8·5	290	1·97	0·050	Wh 48
$RbCr(SO_4)_2.12H_2O$ C.S. 2	290	1·98	$0·082_5$ ±0·003	Ba2 50 a Bl1 50b Ki2 51
	193	1·98 ±0·02	0·063	
	90, 20	1·98 ±0·02	0·054 ±0·001	
$KCr(SeO_4)_2.12H_2O$ C.S. 2	90		0·064 ±0·002	Ba3 u
	20		0·070 ±0·002	
$KAl(SeO_4)_2.12H_2O$ C.S. 2; d. 10	90		$0·089_5$ ±0·001	Ba3 u
	20		0·098 ±0·001	
d. 100	90	1·976 ±0·002	0·0900 ±0·0003	Bl3 51
	20	1·976 ±0·002	0·0983 ±0·0003	

Cr^{3+}* (*contd.*)

Formula	T	g	D	E	A	Remarks	Ref.
$KAl(SeO_4)_2.12D_2O$ C.S. 2	90, 20	1·976 ±0·002	$0·10_5$		$0·0018_5$ ±0·0001		Bl3 51
$K_3[Cr(CN)_6]$ C.S. 7	90, 20	1·99 ±0·01				Only one line	Bo5 52
$K_3[Co(CN)_6]$ C.S. 7; d. 100 to 1000	90 to 12	1·992 ±0·002	+0·083 ±0·001	0·011 ±0·001	0·00147 ±0·00005		† Bo5 52
$K_3[Mn(CN)_6]$ C.S. 7; d. 100	90 to 12	1·993 ±0·002	+0·054 ±0·001	0·012 ±0·001	$0·0014_7$ ±0·0001		‡ Bo5 52
$Al[(CH_3CO)_2CH]_3$ C.S. a; d. 50	290	1·983 ±0·002	0·592 ±0·002	0·052 ±0·002		$M_m=2$, $\psi=+22\frac{1}{2}°$, $a=59°$	§ Si 55

† At 20°K, d.c. of axes:

	a	b	c
z	±0·994	0·104	0
x	±0·104	0·994	0
y	0	0	1

‡ At 90°K, d.c. of axes:

	a	b	c
z	1	0	0
x	0	±0·996	0·087
y	0	±0·087	0·996

* Measurements on some chromic alums have been duplicated by several different authors. Selected values are given in this table, with the most recent publication of any one group of authors.

a, § $Al[(CH_3CO)_2CH]_3$. Astbury (1926): monoclinic, $M=4$. The surroundings of the Al are probably three chelate groups arranged octahedrally. The four magnetic complexes are equivalent in pairs; description of axes as for C.S. 1. The results of Ba2 u, with $d\sim10$, show that there is a gradual transition between 290°K and 90°K. Each of the two sets of complexes is subdivided into three types at 90°K, with either (*a*) slightly differently oriented z axes, or (*b*) slightly different splittings. No further change at 20°K.

Paramagnetic resonance absorption has also been observed in the following compounds at room temperature; (apparent) g-values are given in brackets: $CrBr_3$ (1·99), Ti1 51; $Cr(C_5H_7O_2)_3.H_2O$, Ba1 48; $Cr(C_5H_5N)_2(OH)_2(H_2O)_2Cl$ (1·99), La 51; $Cr(C_7H_8O_2)_3.H_2O$ (2·07), La 51; $Cr[(CH_3)_2(CH)_2CH_2(CO)_2]_3$, $Cr[(C_4H_8)(CO)_2CH.CH_3]_3$, both $g\sim2$, $D>\sim0·5$, Si 55; $CrCl_3$ (1·99), Ko 52; CrF_3 (2·00), Ti1 51; $[Cr(H_2O)_4Cl_2]$ Cl (1·95), La 51; $[Cr(H_2O)_4Cl_2]$ Cl.$2H_2O$, Ko 52; $[Cr(H_2O)_5Cl]$ $Cl_2.H_2O$, Ko 52; $[Cr(NH_3)_5Cl]$ Cl_2 (1·97), $[Cr(NH_3)_6]$ $Cl_3.H_2O$ (1·97), $[Cr(NH_3)_6]$ $(NO_3)_3.H_2O$ (1·95), all La 51; $[Cr(NH_2CH_2)_6]_2(SO_4)_3$, ($g\sim2$, $D\sim0·15$), Si 55; $Cr(NO_3)_3.9H_2O$ (2·26), La 51; $Cr(OH)_3$, Ko 52; $Cr(OH)_3.2H_2O$ (2·00), La 51; $CrPO_4.3H_2O$, Ba1 48; $[Cr(SCN)_2(C_2H_4(NH_2)_2)_2](SCN)$ (1·98), La 51, 52; Cr_2O_3, Ko 52; $Cr_2(SO_3)_3$ (2·00), Ti1 51; $Cr_2(SO_4)_3$ (1·98), Ti1 51; $Cr_2(SO_4)_3.5H_2O$, Ba1 48; $Cr_2(SO_4)_3.15H_2O$, Ba1 48; $Cr_2(SO_4)_3.18H_2O$ (2·00), La 51; $K_3Cr(C_2O_4)_3.3H_2O$ ($g\sim2$, $D\sim0·4$), Si 55.

ground state has only twofold spin degeneracy (see § 2.1), i.e. $S'=\frac{1}{2}$, and there are low-lying levels within a few hundred degrees of the ground state (cf. figure 2.4); at high temperatures rapid spin–lattice relaxation broadens the lines and paramagnetic resonance can thus only be observed at low temperatures. The two cases $S'=\frac{5}{2}$ and $S'=\frac{1}{2}$ are considered separately below.

I. $S'=\frac{5}{2}$.

We shall consider first the ferric compounds, since for these no hyperfine structure has been observed, and accordingly nuclear interactions may be neglected.

(*a*) Fe^{3+} (Table overleaf).

Only some alums have been investigated in detail. The results are interpreted in terms of the following spin Hamiltonian, with $S=\frac{5}{2}$:

$$\mathcal{H}=g\beta(H_xS_x+H_xS_x+H_yS_y)+\tfrac{1}{6}a(S_\xi^4+S_\eta^4+S_\zeta^4-\tfrac{707}{16})+D(S_z^2-\tfrac{35}{12})+\tfrac{7}{36}F(S_z^4-\tfrac{95}{14}S_z^2+\tfrac{81}{16}).$$

Here (ξ, η, ζ) are a set of mutually perpendicular axes with respect to which the z axis is the [111] direction. For more details concerning the terms in a and F, which are only required if $S'>\frac{3}{2}$, see for instance Bleaney and Trenam (1954). With $H\|z$ and $g\beta H\gg a$, the levels are at

$$\begin{aligned}&\pm\tfrac{5}{2}g\beta H+\tfrac{10}{3}D-\tfrac{1}{3}(a-F)\\&\pm\tfrac{3}{2}g\beta H-\tfrac{2}{3}D+(a-F)\\&\pm\tfrac{1}{2}g\beta H-\tfrac{8}{3}D-\tfrac{2}{3}(a-F).\end{aligned}$$

For $H\|x, y$, D is multiplied by $-\frac{1}{2}$, both a and F are multiplied by $\frac{3}{8}$, and the further condition $g\beta H\gg D, F$ is required.

In zero field there are three doublets at:

$$-\tfrac{3}{2}D+(a-F),\ \tfrac{1}{3}D-\tfrac{1}{2}(a-F)\pm\tfrac{1}{6}(\{18D+a-F\}^2+80a^2)^{1/2}.$$

For $H\|z$ the susceptibility χ_z is

$$\frac{35}{32}\frac{g^2}{T}\left(1-\frac{32}{15}\frac{D}{T}\right)\quad (D \text{ in } {}^\circ\text{K}).$$

For a single crystal of an alum the $1/T^2$ term averages to zero.
The value of CT^2/R is

$$\frac{56}{9}D^2+2a^2-\frac{14}{27}F(2a-F)\quad (D, a, F \text{ in } {}^\circ\text{K}).$$

The value of D is sensitive to temperature and dilution.

(*b*) Mn^{2+}.

For crystals with trigonal symmetry (e.g. fluosilicate) the same spin Hamiltonian as for the ferric alums is used, with the additional term

$$A(S_xI_x+S_xI_x+S_yI_y),\quad I=\tfrac{5}{2}.$$

Some authors give g and A to be anisotropic, but since the suggested anisotropy is probably not greater than the maximum expected error, we shall neglect it. The effect of the term in A on the strong field levels (with further condition $g\beta H\gg A$) is to add to the first expression $\pm\frac{5}{2}Am$, to the second $\pm\frac{3}{2}Am$, and to the third $\pm\frac{1}{2}Am$, where m takes the values $\frac{5}{2}, \frac{3}{2}, \ldots, -\frac{5}{2}$. To obtain an indication of the

Fe^{3+}

Formula	T	g	D	a	F	Remarks	Ref.
$KAl(SeO_4)_2.12H_2O$ C.S. 2; d. 300	90	2·003 ±0·003	−0·0103 ±0·0001	−0·0127 ±0·0002	−0·0002 ±0·0002	$\phi=(10{\cdot}5\pm0{\cdot}5)°$	Bl15 54
	20	2·003 ±0·001	−0·0115 ±0·0001	−0·0127 ±0·0001	−0·0002 ±0·0001		
$(NH_3CH_3)Al(SO_4)_2.12H_2O$ C.S. 2; d. 200	90		(−)0·188 ±0·014	(−)0·010 ±0·004			Bl15 54
$NH_4Al(SO_4)_2.12H_2O$ C.S. 2; d. 80	4		0·016 ±0·001	(−)0·0128 ±0·0004			Ub2 51 Me 51
$RbAl(SO_4)_2.12H_2O$ C.S. 2; d. 300	90	2·003 ±0·003	+0·0022 ±0·0002	−0·0134 ±0·0002	−0·0003 ±0·0002	$\phi=(7{\cdot}5\pm0{\cdot}5)°$	Bl15 54
	20	2·003 ±0·001	+0·0031 ±0·0001	−0·0134 ±0·0001	−0·0003$_3$ ±0·0001		

Paramagnetic resonance absorption has also been observed in the following compounds (room temperature, unless stated otherwise); the (apparent) g-values are given in brackets: $Fe(C_5H_7O_2)_3$ (1·95), La 51; $Fe(C_7H_5O_2)_3$, Ba1 48; $Fe(C_{17}H_{33}COO)_3$ (2·00), Ab3 53a; $FeCl(C_6H_4)_6(CN)_8$, T=270–20, (3·8, 2·0), In2 54; $FeCl_3$, Ba1 48; $Fe(NH_4)_3(C_2H_5O_7)_2.2H_2O$ (1·98), La 51; $FeOH(C_2H_3O_2)_2$, $Fe(PO_4).4H_2O$, $Fe_2(C_2O_4)_3$, $Fe_2[C_3H_5(OH)_2OPO_3]_3$, all Ba1 48; $(FeF_3)_3.9H_2O$ (2·02), La 51; $Fe_2(SO_4)_3.3H_2O$ (2·01), Ti1 51; $Fe_2(SO_4)_3.9H_2O$ (2·01), La 51; $Fe_3(CH_3COO)_6(OH)_2NO_3.6H_2O$, T=15 (2·0), Ge u; $Fe_4[Fe(CN)_6]_3$, Ba1 48.

Mn^{2+}

Formula	T	g	D	E	a	A	Remarks	Ref.
$(NH_4)_2Mg(SO_4)_2.6H_2O$ C.S. 1;	290		0·0231 ±0·0002	0·006	0·0003	0·0090 ±0·0002	$\psi=+60°$; $\alpha=30°$	In1 53a
d. 250	290		0·022$_0$	0·004$_1$		0·0095	$\psi=+59°$; $\alpha=30\frac{1}{2}°$	Br 53

$(NH_4)_2Zn(SO_4)_2.6H_2O$ C.S. 1;	290	2·000 ±0·005	0·0238	$0·007_5$	0·0005	0·0091		Ha2 53
d. 1000	230		+0·0243 ±0·0005	0·010 ±0·002	+0·0005 ±0·0001	$-0·0091_1$ ±0·0001	$\psi=+58°$; $\alpha=32°$	Bl7 51a
	195		+0·0258 ±0·0005	$0·008$ $\pm0·001_6$	+0·0007 ±0·0001	$-0·0089_0$ ±0·0001		
	90		+0·0275 ±0·0005	$0·007$ $\pm0·001_4$	+0·0007 ±0·0001	$-0·0089_9$ ±0·0001		
	20	2·00	+0·0277 ±0·0005	0·005 ±0·001	+0·0008 ±0·0002	−0·0093		
$Mg_3Bi_2(NO_3)_{12}.24H_2O$ C.S. 3; d. 200	90	1·99 ±0·02	−0·0211 ±0·0001	0	+0·0008 ±0·0001	−0·0090	Magnetic complex I	Tr 53
	20	1·997 ±0·003	−0·0215 ±0·0001	0	+0·0008 ±0·0001	−0·0090		
	90	1·99 ±0·02	−0·0064 ±0·0001	0	+0·0010 ±0·0001	−0·0089	Magnetic complex II	
	20	1·997 ±0·003	−0·0080 ±0·0001	0	+0·0010 ±0·0001	−0·0090		
$MgSiF_6.6H_2O$ C.S. 4; d. 20 to 150	290		(−)0·0274	0·0030	(+)0·0007	(−)0·0092	$M_m=6$	Ar 54
$ZnSiF_6.6H_2O$ C.S. 4;	290	2·000 ±0·005	0·0171	0	$0·0007_5$	0·0090		Ha2 53
d. 1000	290		−0·0179 ±0·0003	0	+0·0007 ±0·0001	−0·0095		Bl7 51a
	195		−0·0161 ±0·0003	0	+0·0010 ±0·0002	−0·0092		
	90	2·000 ±0·001	−0·0141 ±0·0003	0	+0·0011 ±0·0002	−0·0092		
	20		−0·0134 ±0·0003	0	+0·0009 ±0·0002	−0·0091		

Mn^{3+} (*contd.*)

Formula	T	g	D	E	a	A	Remarks	Ref.
$MgSO_4.7H_2O$ C.S. 6	290	2·000 ±0·005	$0{\cdot}040_0$	~0		0·0088	d.c. of z axes (± ∓0·282, ±0·952, 0·122)	Ha2 53
$CaCO_3$ C.S. a; d. 2000	290	2·002	0·0075	0		0·00878	$F=0{\cdot}0058$	Hu1 54
ZnS p hexag.; d. 10^3 to 10^5	290	2·0024 ±0·0004	0·002			0·0063		Sc 51
	290		0·001			0·0065 ±0·0001		He 52
$Zn(HCOO)_2.2H_2O$	290	$1{\cdot}999_5$ ±0·001	0·0485 ±0·0005	0·011	$0{\cdot}0009_5$	0·0091 ±0·0001	Monoclinic, $M_m=2$ $\psi=+97°$; $a=62°$ (for ψ, a, see C.S. 1)	In1 53b
$Zn(CH_3COO)_2.3H_2O$	290		$0{\cdot}0235$	$0{\cdot}002_5$		0·0084	Monoclinic, $M_m=1$ $z\equiv c$, $x\equiv b$	Ku2 52 Ha2 53
$Zn(CH_3COO)_2.4H_2O$ d. 100	290	2·00 ±0·01	0·0412	$0{\cdot}006_6$	0·0008	0·0087	Monoclinic, $M_m=2$ $\psi=+47°$, $a=29°$ (for ψ, a, see C.S. 1)	In1 53a

a $CaCO_3$ calcite. Wyckoff (1920): hexagonal, $M=2$. The nearest neighbours to each Ca are six O, with trigonal symmetry. With Mn, the two magnetic complexes are completely equivalent.

Paramagnetic resonance absorption has also been observed in the following (room temperature):

(*a*) Phosphors containing small concentrations of Mn (He 52); in each case g is close to 2·0: cubic ZnS ($A=0{\cdot}0064$); $ZnAl_2O_4$ ($A=0{\cdot}0076_5$); MgO [197–2] ($A=0{\cdot}0082$); MgO [200–2] ($A=0{\cdot}0072$); $CdSiO_3$; Mg_2GeO_4; ZnF_2; ZnS: Tb; Zn_2SiO_4; 8ZnO.BeO.5SiO_2; Zn_2GeO_4; $Zn_3(PO_4)_2$.

(*b*) Undiluted compounds; (apparent) g-values given in brackets: $Mn(BO_2)_2$ (2·01), La 51; $Mn(CH_3COO)_2.2H_2O$ (2·00), Ko 52; $Mn(C_3H_5O_2)_2.3H_2O$, Ba1 48; $Mn(C_6H_4)_4(CN)_8$ (2·0), T down to 20, In2 54; $Mn(C_{15}H_{31}COO)_2$ (2·00), Ab3 53a; $Mn(C_{17}H_{35}COO)_2$ (2·00), Ab3 53a; $MnCO_3$ (2·00), La 51; $MnC_2O_4.2H_2O$ (2·00), La 51; $MnCl_2.4H_2O$ (2·00), La 51 and (2·09), Ko 52; MnF_2, Ba1 48; $Mn(H_2PO_2)_2.H_2O$, Ba1 48; $MnK_2(SO_4)_2.6H_2O$, Ba1 48; $MnSO_4$ (2·00), Ko 52, Ku1 51; $MnSO_4.H_2O$ (2·00) Ku1 51; $MnSO_4.4H_2O$ (2·00–2·07), Ku1 51; $MnSO_4.5H_2O$ (2·06) Ku1 51; $Mn_2Fe(CN)_6.7H_2O$, $Mn_2(P_2O_7).3H_2O$, $Mn_3(AsO_4)_2$, $Mn_3(C_6H_5O_7)_2$, $Mn_3(PO_4)_2.7H_2O$, all Ba1 48.

spread of the levels in zero field due to A, let us take $D \gg a$, F, A; then if D is positive the top levels are spread over $25A/2$, the middle ones over $15A/2$, and the lowest over $9A$; this order is reversed if D is negative. The susceptibility is as for Fe^{3+}, but the expression for CT^2/R requires the additional term $1225A^2/48$ (A in °K).

On the other hand, often the symmetry of the magnetic complex is not higher than orthorhombic, as in the Tutton salts; the following spin Hamiltonian is then used, with $S=\frac{5}{2}$ and $I=\frac{5}{2}$:

$$\mathcal{H}=g\beta(H_zS_z+H_xS_x+H_yS_y)+\tfrac{1}{6}a(S_z^4+S_x^4+S_y^4-\tfrac{707}{16})+D(S_z^2-\tfrac{35}{12})$$
$$+E(S_x^2-S_y^2)+A(S_zI_z+S_xI_x+S_yI_y).$$

With $H \| z$ and $g\beta H \gg E$, a, A, the levels are at

$$\pm\tfrac{5}{2}g\beta H+\tfrac{10}{3}D+\tfrac{1}{2}a\pm\tfrac{5}{2}Am$$
$$\pm\tfrac{3}{2}g\beta H-\tfrac{8}{3}D-\tfrac{3}{2}a\pm\tfrac{3}{2}Am \quad (m=\tfrac{5}{2},\ \tfrac{3}{2},\ \ldots,\ -\tfrac{5}{2})$$
$$\pm\tfrac{1}{2}g\beta H-\tfrac{8}{3}D+a\pm\tfrac{1}{2}Am.$$

With $H \| x, y$, the further condition $g\beta H \gg D$ is required; D is multiplied by $-\frac{1}{2}$, and the term in a is left unaltered. Neglecting the nuclear interaction, there are three doublets in zero field; neglecting a, these are at $(4\sqrt{7}/3)(1+k)^{1/2}D\cos\phi$, where $\cos 3\phi=-10(1-3k)[7(1+k)]^{-3/2}$, $k=3E^2/D^2$. The remarks made above for the trigonal case about the effect of A on the zero field levels also apply to the rhombic case.

The susceptibility is as for Fe^{3+}, but the expression for CT^2/R becomes

$$\frac{56}{9}(D^2+3E^2)+2a^2+\frac{1225}{48}A^2 \quad (D, E, a, A \text{ in } °K).$$

Only D and E are expected to be sensitive to temperature and dilution.

In the tables, where authors give g or A to be anisotropic, the mean of their anisotropic parameters are given.

II. $S'=\frac{1}{2}$.

For the cyanides, $S'=\frac{1}{2}$ (see § 2.1). The nuclear interactions can be neglected for the ferricyanide, and the results are interpreted in terms of the following spin Hamiltonian, with $S=\frac{1}{2}$:

$$\mathcal{H}=\beta(g_zH_zS_z+g_xH_xS_x+g_yH_yS_y).$$

With $H \| z$, the levels are at $\pm\frac{1}{2}g_z\beta H$.

Fe^{III}

Formula	T	g			Ref.
$K_3[Fe(CN)_6]$ C.S.7	20	g_a 2·30 ±0·03	g_b 2·18 ±0·03	g_c 0·94 ±0·03	Bl7 52
$K_3[Co(CN)_6]$ C.S.7; d. 100	20	g_z 0·91$_5$ ±0·01	g_x 2·35 ±0·02	g_y 2·10 ±0·02	†Ba5 55

† d.c. of axes:		a	b	c
	z	0	0	1
	x	±0·866	0·500	0
	y	±0·500	0·866	0

For the manganocyanide, the term $A_xS_xI_x+A_xS_xI_x+A_yS_yI_y$ must be added to the spin Hamiltonian. With $H\|z$ and $g\beta H \gg A$, the levels are at $\pm\frac{1}{2}g_x\beta H\pm\frac{1}{2}A_xm$ ($m=\frac{5}{2}, \frac{3}{2}, \ldots, -\frac{5}{2}$).

In zero field, making the simplifying assumption that $A_x=A_z=C$, there are two singlets at $-\frac{1}{4}A_y\pm\frac{3}{2}C$, and five doublets at $\frac{5}{4}A_y$, $-\frac{1}{4}A_y\pm\frac{1}{2}(4A_y^2+5C^2)^{1/2}$, $-\frac{1}{4}A_y\pm\frac{1}{2}(A_y^2+8C^2)^{1/2}$. The value of CT^2/R is $\frac{35}{48}(A_x^2+A_y^2+A_z^2)$ (A in °K). A slight dependence of g and A on temperature and dilution, although not reported, would not be unexpected.

Mn^{II}

$K_4[Fe(CN)_6]3H_2O$ C.S.7; d. 100	$T=15$	g_z	0·72 ±0·05	g_x	2·624 ±0·008	g_y 2·182 ±0·008
		A_z	0·0083 ±0·0013	A_x	0·00845 ±0·00005	A_y 0·00465 ±0·00005

		a	*b*	*c*	
d.c. of axes:	*z*	0·295	±0·899	−0·322	
	x	0·864	±0·105	0·495	
	y	0·410	±0·423	−0·807	Ref. Ba5 55

3d^6 *Iron* Fe^{2+}.

A cubic (octahedral) field leaves an orbital triplet lowest, with fivefold spin degeneracy. Since the number of electrons is even, a rhombic field lifts all degeneracy. Only one crystal (fluoride) has been examined in detail, and in this crystal the ground state consists of a slightly split doublet. Low temperatures are required to observe the spectrum, since at high temperatures rapid spin–lattice relaxation broadens the lines. The results are interpreted in terms of the following spin Hamiltonian, with $S=\frac{1}{2}$, where $g_\perp$ has been taken to be zero:

$$\mathcal{H}=g_z\beta H_zS_z+\Delta S_x.$$

The energy levels are at $\pm\frac{1}{2}(g_z^2\beta^2H^2\cos^2\theta+\Delta^2)^{1/2}$ where θ is the angle H makes with z.

The value of CT^2/R is $\frac{1}{4}\Delta^2$ (Δ in °K).

	T	g_z	Δ	Ref.
ZnF_2; d. ~3000	90	8·97±0·05	0·224+0·010	Ti2 55
	20	8·97±0·02	0·203±0·004	

C.S. Ferrari (1926): tetragonal, $M=2$. Each Zn is surrounded by a distorted octahedron of six F. The two magnetic complexes in unit cell are equivalent; the spectrum has axial symmetry about the crystal *c*-axis.

Paramagnetic resonance absorption has also been observed in:

$K_2Fe(SO_4)_2.6H_2O$, $(NH_4)_2Fe(SO_4)_2.6H_2O$, $FeSiF_6.6H_2O$ (all $T=20$), quoted in I.

3d^7 *Cobalt* Co^{2+}.

I. $S'=\frac{1}{2}$.

A cubic (octahedral) field leaves an orbital triplet lowest; this triplet, with its fourfold spin degeneracy, is split into Kramers doublets by the combined effect of fields of lower symmetry and spin–orbit coupling (cf. figure 2.3). Low temperatures are generally required to observe paramagnetic resonance, since at high temperatures rapid spin–lattice relaxation broadens the lines. The results on the

Co^{2+} ($S'=\frac{1}{2}$)

Formula	T	g			A			Remarks	Ref.
$K_2Zn(SO_4)_2.6H_2O$ C.S. 1; d. 500 to 10^6	20	g_z 6·56 ±0·13	g_{min} 2·50 (K_1K_3) ±0·05	g_2 3·35 ±0·07	A_z 0·0286 ±0·0006	A_{min} 0·0065 (K_1K_3) ±0·0003	A_2 0·0080 ±0·0004	$\psi=+163°$; $\alpha=35°$	Bl7 51b
$(NH_4)_2Zn(SO_4)_2.6H_2O$ C.S. 1; d. 500 to 10^6	20	g_z 6·45 ±0·13	g_{min} 3·06 (K_1K_3) ±0·06	3·06 g_2 ±0·06	A_z 0·0245 ±0·0005	A_{min} 0·0020 (K_1K_3) ±0·0001	A_2 0·0020 ±0·0001	$\psi=+130°$; $\alpha=34°$	Bl7 51b
$Rb_2Zn(SO_4)_2.6H_2O$ C.S. 1	20	g_z 6·6$_5$	g_{min} 2·7 (K_1K_3)	g_2 3·3	A_z 0·0293 ±0·0003	A_{min} 0·0049 (K_1K_3) ±0·0005		$\psi=+157°$; $\alpha=37°$	Bo3 u
$Mg_3Bi_2(NO_3)_{12}.24H_2O$ C.S. 3; d. 100	20	$g_\parallel$ 7·29 ±0·01	$g_\perp$ 2·338 ±0·004		A 0·0283 ±0·0001	$B\leqslant$0·0001		Magnetic complex I	Tr 53
		$g_\parallel$ 4·108 ±0·003	$g_\perp$ 4·385 ±0·003		A 0·0085 ±0·0001	B 0·0103 ±0·0001		Magnetic complex II	
$ZnSiF_6.6H_2O$ C.S. 4; d. 500 to 10^6	20	$g_\parallel$ 5·82 ±0·12	$g_\perp$ 3·44 ±0·07		A 0·0184 ±0·0004	B 0·0047 ±0·0002		Main line; $g_\parallel$ axis $\parallel$ c-axis of crystal	Bl7 51b
		6·6 ±0·1	2·62 ±0·05		0·023	0·0009		Much weaker lines with different axes (defective lattice points)	
		6·6 ±0·2	2·82 ±0·06		0·025	0·0013			
		3·58 ±0·05	4·09 ±0·05		0·01	0·01			
$ZnSO_4.7H_2O$ C.S. 6; d. 100	20	g_x 6·90 ±0·14	g_z 2·30 ±0·05	g_y 3·30 ±0·07	A_x 0·0254 ±0·0005	A_z 0·0028 ±0·0001		†	Bl7 51b

† For one complex: z in (110) with $\angle zOc=13°$; x in (110) with $\angle xOc=103°$; $y\parallel$ [110]. Other axes are given by reflection symmetry. Undiluted $CoSO_4.7H_2O$ has also been examined: g is anisotropic, e.g. in ab plane from 1·4 to 5·8.

Paramagnetic resonance absorption has also been observed in $Co(C_6H_4)_4(CN)_8$, $T=270$–20; two complexes: I $g=2·9$, II $g=2·4-1·98$; ref. In2 54.

hydrated cobalt crystals (see table p. 347) are interpreted in terms of the following spin Hamiltonian with $S=\frac{1}{2}$ and $I=\frac{7}{2}$:

$$\mathcal{H}=\beta(g_zH_zS_z+g_xH_xS_x+g_yH_yS_y)+A_zS_zI_z+A_xS_xI_x+A_yS_yI_y.$$

For some crystals the spectrum has axial symmetry, i.e. $g_x=g_y=g_\perp$, $A_x=A_y=B$. With $H\|z$ and $g\beta H\gg A$, the levels are at $\pm\frac{1}{2}g_z\beta H\pm\frac{1}{2}A_zm$, $(m=\frac{7}{2}, \frac{5}{2}, \ldots, -\frac{7}{2})$. Simple formulae for the energy levels in zero field can only be given if there is axial symmetry; there are then two singlets at $-\frac{1}{2}A_z\pm 2B$, and seven doublets at $\frac{7}{4}A_z$, $-\frac{1}{4}A_z\pm\frac{1}{2}(9A_z^2+7B^2)^{1/2}$, $-\frac{1}{4}A_z\pm(A_z^2+3B^2)^{1/2}$, $-\frac{1}{4}A_z\pm\frac{1}{2}(A_z^2+15B^2)^{1/2}$. With rhombic symmetry, the doublets are split.

The value of CT^2/R is $\frac{21}{16}(A_x^2+A_y^2+A_z^2)$ (A in °K). A slight dependence of g and A on temperature and dilution, although not reported, would not be unexpected.

II. $S'=\frac{3}{2}$.

In Cs_3CoCl_5 the magnetic complex is tetrahedral. This gives a cubic field of opposite sign to that in the octahedral complexes, and leaves an orbital singlet lowest (cf. $3d^3$); the spin quadruplet is split into two Kramers doublets by fields of lower symmetry. The results are interpreted in terms of the following spin Hamiltonian, with $S=\frac{3}{2}$:

$$\mathcal{H}=g_\parallel\beta H_zS_z+g_\perp\beta(H_xS_x+H_yS_y)+D(S_z^2-\tfrac{5}{4}).$$

The nuclear interactions have here been omitted since, owing to the large line width, the hyperfine structure has not been observed. With $H\|z$, the levels are at $\pm\frac{3}{2}g_\parallel\beta H+D$, $\pm\frac{1}{2}g_\parallel\beta H-D$. With $H\perp z$, the levels are at

$$\tfrac{1}{2}g_\perp\beta H\pm(g_\perp^2\beta^2H^2-g_\perp\beta HD+D^2)^{1/2},\quad -\tfrac{1}{2}g_\perp\beta H\pm(g_\perp^2\beta^2H^2+g_\perp\beta HD+D^2)^{1/2}.$$

With $H\|z$, the susceptibility χ_z is

$$\frac{15}{32}\frac{g_\parallel^2}{T}\left(1-\frac{4D}{5T}\right)\quad (D \text{ in } °K).$$

The value of CT^2/R is D^2 (D in °K).

	T	$g_\parallel$	$g_\perp$	D	E	Ref.
Cs_3CoCl_5	90, 20	2·32 ±0·04	2·27 ±0·04	~−4·5	0	Ow 55b

C.S. Powell and Wells (1935): tetragonal, M=4. Each Co is surrounded by a distorted tetrahedron of 4Cl. The four magnetic complexes in unit cell are equivalent; the spectrum has axial symmetry $\|c$.

$3d^8$ *Nickel* Ni^{2+}.

A cubic (octahedral) field leaves an orbital singlet lowest; the spin triplet is split into a doublet and a singlet by a trigonal or tetragonal field, and into three singlets by a rhombic field. In the crystals which have been examined (see table below) the hyperfine structure of the low abundant odd isotope has not been observed, and so the nuclear interactions will be omitted. The results are interpreted in terms of the following spin Hamiltonian, with $S=1$:

$$\mathcal{H}=\beta(g_zH_zS_z+g_xH_xS_x+g_yH_yS_y)+D(S_z^2-\tfrac{2}{3})+E(S_x^2-S_y^2).$$

A simplification is that the resonance results show g to be very nearly isotropic. Furthermore, $E=0$ for some crystals. With $H\|z$, the levels are $-\frac{2}{3}D$, $+\frac{1}{3}D\pm(E^2+g_z{}^2\beta^2H^2)^{1/2}$. For $H\|z$, the susceptibility χ_z is

$$\frac{g_z{}^2}{4T}\left(1-\frac{D}{3T}\right) \quad (D \text{ in } {}^\circ\text{K}).$$

A full discussion of the susceptibility of the Tutton salts is given by Griffiths and Owen (1952a). The value of CT^2/R is $\frac{2}{9}(D^2+3E^2)$ (D, E in °K). The values of D and E are sensitive to temperature and dilution.

3d^9 *Copper* Cu^{2+}.

A cubic (octahedral) field leaves a ' nonmagnetic ' (Van Vleck 1932b) orbital doublet lowest. A tetragonal or rhombic field splits this orbital doublet (cf. figure 2.3), leaving two Kramers doublets, and paramagnetic resonance is observed for the lower of these. This is the situation for the Tutton salts which are given in table A below, and also for the compounds in table C, which contains results for undiluted crystals in which the six neighbours of the cupric ion are not all water molecules; in these crystals exchange effects are often important.

A trigonal field, however, does not split the lower orbital doublet. The situation is then more complicated, since the orbital degeneracy is lifted by the Jahn–Teller effect, which distorts the magnetic complex (see Bleaney, Bowers and Trenam 1955). Experimentally, one observes at low temperatures a transition in the spectrum of crystals which have trigonal symmetry. Below the transition temperature, the spectrum corresponds to three magnetic complexes, each with tetragonal symmetry; the three tetragonal axes are the edges of a cube whose trigonal axis is the crystal axis. The results for the trigonal crystals are given in table B, for temperatures both above and below the transition. The susceptibility of the crystal is left practically unaltered by the transition, but the energy levels and the value of CT^2/R are changed.

For the crystals which are reasonably dilute, i.e. for those given in tables A and B, the results are interpreted in terms of the following spin Hamiltonian, with $S=\frac{1}{2}$ and $I=\frac{3}{2}$:

$$\mathcal{H}=\beta(g_xH_xS_x+g_xH_xS_x+g_yH_yS_y)+A_xS_xI_x+A_xS_xI_x+A_yS_yI_y+P(I_z{}^2-\tfrac{5}{4})$$
$$+P'(I_x{}^2-I_y{}^2).$$

For a precise analysis of the spectrum, the additional term $-g_N\beta_N\mathbf{H}.\mathbf{I}$ is required, but it will be omitted here because it is relatively unimportant.

With $H\|z$ and $g\beta H\gg A\gg P'$, the levels are at: $\pm\frac{1}{2}g_z\beta H\pm\frac{1}{2}A_zm+P(m^2-\frac{5}{4})$, ($m=\frac{3}{2}$, $\frac{1}{2}$, $-\frac{1}{2}$, $-\frac{3}{2}$). To apply the transformation rules for $H\|x$, y, the further condition $A\gg P$ is required; this condition is often not fulfilled, in which case the secular equation has to be solved.

Simple formulae for the energy levels in zero field can only be obtained on the assumption of axial symmetry, i.e. $A_x=A_y=B$, $P'=0$. This approximation is quite good for crystals with trigonal symmetry, but rather unsatisfactory for most others. With axial symmetry, the levels are two singlets at $-\frac{1}{2}A_z-P\pm B$, and three doublets at $\frac{3}{4}A_z+P$ and $-\frac{1}{4}A_z\pm\frac{1}{2}\{A_z-2P\}^2+3B^2)^{1/2}$. With $H\|z$, the susceptibility χ_z is $(3/32)g_z{}^2/T$. Care must be exercised in using this formula to calculate the anisotropy in the susceptibility, especially for Tutton salts, since the magnitude and axes of the anisotropy are temperature dependent.

Ni^{3+}

Formula	T	g	D	E	Axes* $\psi(^\circ)$	Axes* $\theta(^\circ)$	Axes* $\alpha(^\circ)$	Remarks	Ref
$K_2Ni(SO_4)_2.6H_2O$ C.S. 1	290	2·25 ±0·05	−3·30	$-0{\cdot}51_7$	−12·5	11	~45		Gr1 52a
$(NH_4)_2Ni(SO_4)_2.6H_2O$ C.S. 1	290	2·25 ±0·05	−2·24	$-0{\cdot}38_7$	−14	3½	45	In $(NH_4)_2Zn(SO_4)_2.6H_2O$ (d. 50) parameters and axes practically unchanged	Gr1 52a
	90	2·25 ±0·05	−1·99	$-0{\cdot}48_6$					
$Tl_2Ni(SO_4)_2.6H_2O$ C.S. 1	290	2·25 ±0·05	−2·6	−0·10	−11	11	45		Gr1 52a
$K_2Ni(SeO_4)_2.6H_2O$ C.S. 1	290	2·25 ±0·05	−3	−1	−13	~0	~45		Gr1 52a
$(NH_4)_2Ni(SeO_4)_2.6H_2O$ C.S. 1	290	2·25 ±0·05	−1·89	−0·79	−28	0	50		Gr1 52a
	90	2·25 ±0·05	−1·73	−0·82					
$Mg_3La_2(NO_3)_{12}.24H_2O$ C.S. 3; d. 800	90	2·24	0·177 ±0·002	0					Tr u
$NiSiF_6.6H_2O$ C.S. 4	290	2·3	−0·50	0					Ho 49
	195	2·29	−0·32	0					Pe1 50
	90	2·26	−0·17	0				In $ZnSiF_6.6H_2O$ (d. 4, 16), D is ~ 20% greater than in undiluted at all T	Gr1 52a
	60		−0·14	0					
	20	2·29	−0·12	0					
	14		−0·12	0					

$Ni(BrO_3)_2.6H_2O$ C.S. 5	290	2·29 ±0·04	1·93 ±0·04	0		Ow1 u
$NiSO_4.7H_2O$ C.S. 6	(290)	$2{\cdot}2_6$	$-3{\cdot}5_6$	$-1{\cdot}5_0$	†	On1 53

* For the two ions, z_1 and z_2 in K_1OK_3 plane, with $\angle z_1OK_1$, $\angle z_2OK_1 = \pm\theta$, where $\angle K_1Oc = \psi$; x_1 is approx. $\| y_2$, both lying nearly in K_2OK_3 plane, with $\angle x_1OK_2 = a$, $\angle y_2OK_2 = 90 - a$.

† For one complex, d.c. are z: (0·95, 0·31, 0), y: (−0·31, 0·95, 0·09). Axes of other complexes are given by reflection symmetry.

Paramagnetic resonance absorption has also been observed in the following compounds, unless stated otherwise at room temperature; the g-values are given in brackets:

$NiBr_2$, (2·27), Ti1 51; $NiBr_2(NH_3)_6$ (2·14), La 51; $NiCl_2$ (2·21), Ti1 51; $NiCl_2.6H_2O$, Ba1 48; $NiI_2(NH_3)_6$, Ow1 u; $Ni(C_6H_4)_4(CN)_8$, $T = 270$ to 20, (2·20), In2 54.

Table A. Cu^{2+} Tutton Salts (C.S. 1)

Formula	T	g			A			P	$\psi(°)$	$\alpha(°)$	Ref.
$Cs_2Cu(SO_4)_2.6H_2O$	90	g_z 2·43 ±0·02	g_{min} (K_1K_3) 2·08 ±0·02	g_2 2·06 ±0·02					+114	40	Bl12 49
$K_2Cu(SO_4)_2.6H_2O$	290	g_1 2·31 ±0·03	g_2 2·07 ±0·03	g_3 2·25 ±0·03					+105		Ba2 52
	90	g_z 2·36 ±0·02	g_x 2·14 ±0·02	g_y 2·04 ±0·02					+105	42	Bl12 49
$K_2Zn(SO_4)_2.6H_2O$ d. 50 to 2000	290	g_z 2·05 ±0·03	g_{max} (K_2K_3) 2·25 ±0·03	g_1 2·26 ±0·03					+15	32	Ba2 52
	90, 20	g_z 2·44 ±0·02	g_{min} (K_1K_3) 2·13 ±0·02		A_z 0·0103 ±0·0005	A_{min} (K_1K_3) 0·0034 ±0·0005		0·0011 ±0·0001	~+105	42	Bl4 55
$K_2Zn(SO_4)_2.6D_2O$ d. 200 to 1000	20	g_z 2·42 ±0·01	g_x 2·16 ±0·01	g_y 2·04 ±0·01	A_z −0·0099 ±0·0001	A_x <0·0017	A_y +0·0061 ±0·0003	+0·00110 ±0·00005 and P' = 0·00011 ±0·00005	~+105	43	† Bl4 55
$K_2Cu(SeO_4)_2.6H_2O$	90	g_z 2·38 ±0·02	g_{min} (K_1K_3) 2·07 ±0·02	g_2 2·04$_5$ ±0·02					+73	37	Bl12 49
$(NH_4)_2Cu(SO_4)_2.6H_2O$	290	g_1 2·32 ±0·03	g_2 2·09 ±0·03	g_3 2·25 ±0·03					+77		Ba2 52
	90	g_z 2·45 ±0·02	g_{min} (K_1K_3) 2·12 ±0·02	g_2 2·06 ±0·02					+65	39	Bl12 49

$(NH_4)_2Zn(SO_4)_2.6H_2O$ d. 50 to 2000	290	g_z 2·04 ±0·03	g_{max} 2·26 (K_2K_3) ±0·03	g_1 2·28 ±0·03						+167	32	Ba2	52
	20	g_z 2·46 ±0·02	g_x 2·12 ±0·02	g_y 2·05 ±0·02	A_z 0·0130 ±0·0005	A_x 0·0025 ±0·0005	A_y 0·0035 ±0·0005	0·0011 ±0·0001		+65	38	Bl4	55
$(NH_4)_2Cu(SeO_4)_2.6H_2O$	90	g_z 2·39 ±0·02	g_{min} 2·07$_5$ (K_1K_3) ±0·02	g_y 2·06$_5$ ±0·02						+72	37½	Bl12	49
$Rb_2Cu(SO_4)_2.6H_2O$	290	g_1 2·28 ±0·03		g_3 2·24 ±0·03						+105		Ba2	52
	90	g_z 2·45 ±0·02	g_{min} 2·11 (K_1K_3) ±0·02	g_2 2·07 ±0·02						+105	40	Bl12	49
$Rb_2Zn(SO_4)_2.6H_2O$ d. 50 to 2000	290	g_z 2·08 ±0·03	g_{max} 2·27 (K_2K_3) ±0·03	g_1 2·25 ±0·03						+15	33	Ba2	52
	20	g_z 2·44 ±0·02	g_{min} 2·12 (K_1K_3) ±0·02		A_z 0·0116 ±0·0005	A_{min} 0·0030 (K_1K_3) ±0·0005		0·0011 ±0·0001	~	+105	42	‡Bl4	55
$Rb_2Zn(SO_4)_2.6D_2O$ d. 200 to 1000	20	g_z 2·43 ±0·02	g_x 2·15 ±0·02	g_y 2·04 ±0·02	A_z −0·0110 ±0·0002	A_x <0·0020	A_y +0·0059 ±0·0004	+0·0012 ±0·0001	~	+105	42	‡Bl4	55
$Tl_2Cu(SO_4)_2.6H_2O$	90	g_z 2·40 ±0·02	g_{min} 2·08 (K_1K_3) ±0·02	g_2 2·06 ±0·02						+112	39½	Bl12	49

† $^{65}A/^{63}A = 1{\cdot}069 \pm 0{\cdot}003$, $^{63}P/^{65}P = 1{\cdot}08 \pm 0{\cdot}02$

‡ As the crystal is warmed up from 20°K, an anomalous second spectrum appears. At 90°K the intensity of the anomalous spectrum is roughly ⅓ of that of the normal spectrum: the g-values of the two spectra differ by less than 1%.

Table B. Cu^{2+} Trigonal Crystals

Formula	T	g		A			Temp. of transition	Ref.
$Mg_3Bi_2(NO_3)_{12}.24H_2O$ C.S. 3; d. 100	90	$g_{\parallel}$ 2·219 ±0·003	$g_{\perp}$ 2·217 ±0·003	A 0·0027 ±0·0001	B 0·0026 ±0·0001			Bl5 55
	20	g_z 2·454 ±0·003	$g_{x,y}$ 2·096 ±0·003	A_z 0·0110 ±0·0001	$A_{x,y}$ 0·0017 ±0·0002			
$Cu_3La_2(NO_3)_{12}.24H_2O$ C.S. 3	290	2·22					173°K to 273°K	Bi 53
	90, 20	g_z 2·41	$g_{x,y}$ 2·10					
$Mg_3La_2(NO_3)_{12}.24D_2O$ C.S. 3; d. 500	90	$g_{\parallel}$ 2·219 ±0·003	$g_{\perp}$ 2·218 ±0·003	A 0·00290 ±0·00005	B 0·00275 ±0·00005		33°K to 45°K	Bl5 55 Bi 53
	45	2·235 ±0·005						
	20	g_z 2·470 ±0·002	$g_{x,y}$ 2·097 ±0·002	A_z −0·0113 ±0·00005	A_x +0·00190 ±0·00005	A_y +0·00123 ±0·00005		† Bl5 55
$CuSiF_6.6H_2O$ C.S. 4	(290)	2·20						Yo 54
$ZnSiF_6.6H_2O$ C.S. 4	90	$g_{\parallel}$ 2·221 ±0·005	$g_{\perp}$ 2·230 ±0·005	A 0·0021 ±0·0005	B 0·0028 ±0·0005		12°K to 50°K	Bl5 55
	20, 12	g_z 2·46 ±0·01	$g_{x,y}$ 2·10 ±0·01	A_z 0·0110 ±0·0003	$A_{x,y}$ <0·0030			
$Zn(BrO_3)_2.6H_2O$ C.S. 5; d. 100	90	2·21$_7$ ±0·01			0·0028 ±0·0005		<7°K to 35°K	Bl5 55

† $P = +0{\cdot}00111 \pm 0{\cdot}00005$, $P' = -0{\cdot}00004 \pm 0{\cdot}00001$

The value of CT^2/R is $\frac{5}{16}(A_x^2+A_y^2+A_z^2)+P^2+3P'^2$ (A, P, P' in °K). The tabulated values are the mean of the values for the two isotopes ^{63}Cu and ^{65}Cu, weighted in proportion to their natural abundances. The hyperfine structure cannot normally be resolved at room temperature, since rapid spin–lattice relaxation broadens the lines too much.

Exchange effects.

A number of cupric crystals show, in different ways, interesting exchange effects. (It may be remarked that the g-values remain similar to those of the crystals discussed above.) Three examples are, in order of increasing exchange energy, $CuSO_4.5H_2O$, $CuCl_2.2H_2O$ and $Cu(CH_3COO)_2.H_2O$. For the sulphate, the exchange energy between neighbouring magnetic complexes is of order 0·15 cm^{-1}. This is of the order of a microwave quantum, and the details of the spectrum depend on the microwave frequency used; $K_2CuCl_4.2H_2O$ and $(NH_4)_2CuCl_4.2H_2O$ are similar. Cupric chloride becomes antiferromagnetic below 4·3°K, with a change in the spectrum. In the acetate, there is relatively strong (~300 cm^{-1}) exchange coupling between isolated pairs of copper ions, forming a lower singlet state and an upper triplet state, the latter only being paramagnetic; the results for this paramagnetic level (about 440°K above the ground state) are interpreted in terms of the following spin Hamiltonian, with $S=1$:

$$\mathcal{H}=\beta(g_zH_zS_z+g_xH_xS_x+g_yH_yS_y)+D(S_z^2-\tfrac{2}{3})+E(S_x^2-S_y^2)+\mathcal{H}_N.$$

$\mathcal{H}_N$ refers to the nuclear interactions (see Bleaney and Bowers 1952). Copper propionate is similar to the acetate. (For table see overleaf.)

5.4. *The* 4d *and* 5d *Groups*

These elements generally form covalent complexes, of which a considerable number are diamagnetic, e.g. $[Mo^{iv}(CN)_8]^{4-}$, $4d^2$; $[Pd^{ii}Cl_4]^{2-}$, $4d^8$; $[Pt^{iv}Cl_6]^{2-}$, $5d^6$. In many of the paramagnetic salts, especially those of the 5d group, there appears to be particularly strong interaction between paramagnetic neighbours; in such cases, to observe the resonance spectrum it is essential to use magnetically dilute mixed crystals, as otherwise the lines are too broad. The number of salts which are suitable for resonance measurements is rather limited, and the available data are correspondingly fewer than for salts of the 3d and 4f groups (for a survey see Griffiths, Owen and Ward 1953).

The ions for which resonance has been observed, together with the free ion spin S, the effective spin S' and g values, are given in table 5.3.

Table 5.3

Configuration	Ion	S	S'	g	r.m.s. g
$4d^1$	Mo^{5+}	$\frac{1}{2}$	$\frac{1}{2}$	1·94—1·96	1·95
$4d^3$	Mo^{3+}	$\frac{3}{2}$	$\frac{3}{2}$	1·90—2·0	1·95
$4d^5$	Ru^{3+}	$\frac{5}{2}$	$\frac{1}{2}$	1·0 —3·24	1·9
$4d^9$	Ag^{2+}	$\frac{1}{2}$	$\frac{1}{2}$	2·04—2·18	2·1
$5d^3$	Re^{4+}	$\frac{3}{2}$	$\frac{3}{2}$		~1·8
$5d^5$	Ir^{4+}	$\frac{5}{2}$	$\frac{1}{2}$	0·75—2·20	1·8

Table C. Cu^{2+}

Formula	T	g			Remarks	Ref.
$Cu(NH_3)_4(NO_3)_2$	290	[1] g_a 2·07 $\pm$0·02	g_b 2·14 $\pm$0·02	g_c 2·02 $\pm$0·02	orthorhombic (?) (Groth 1908)	Ok 54
$Cu(NH_3)_4SO_4.H_2O$	290	g_a 2·02 $\pm$0·02	g_b 2·05 $\pm$0·02	g_c 2·15 $\pm$0·02	orthorhombic (Groth 1908)	Ok 54
$Cu[(CH_3CO)_2CH]_2$ C.S. a	290, 90	$g_\parallel$ 2·28 $\pm$0·03	$g_\perp$ 2·07 $\pm$0·03		$M_m=2$	Ba1 u
$CuCl_2.2H_2O$ C.S. b	290 to 15	g_a 2·187 $\pm$0·005	g_b 2·037 $\pm$0·005	g_c 2·252 $\pm$0·005	only one line observed	Ge u
	4·2 to 1·2				Antiferromagnetic below 4·3°K; spectrum described by Ub1 52	
$CuSO_4.5H_2O$ C.S. c	290	$g_\parallel$ 2·08 g_x 2·46	$g_\perp$ 2·27 $g_{x,y}$ 2·08		One line, $\tilde{\nu}\leqslant 0{\cdot}8$ cm^{-1} Two lines, $\tilde{\nu}=1{\cdot}2$ cm^{-1} See text	† Ba2 50b
$K_2CuCl_4.2H_2O$ $(NH_4)_2CuCl_4.2H_2O$ C.S. d	290	$g_{\parallel(c)}$ 2·06 $g_x>2{\cdot}30$	$g_\perp$ 2·22 $g_x<2{\cdot}10_5$	$g_{y(c)}$ 2·06	One line, $\tilde{\nu}\leqslant 1{\cdot}5$ cm^{-1} Two lines, $\tilde{\nu}=1{\cdot}8_5$ cm^{-1} See text	Ab2 54
$Cu(CH_3COO)_2.H_2O$ C.S. e	290 90	g_z 2·34$_4$ $\pm$0·01	g_x 2·053 $\pm$0·005	g_y 2·093 $\pm$0·005	$D=0{\cdot}345\pm0{\cdot}005$, $E=0{\cdot}007\pm0{\cdot}003$; $A_z=0{\cdot}008$, $A_{x,y}<0{\cdot}0010$; $\psi=-34°$, $\alpha=33°$; $S=1$, see text	Ab3 53b Bl3 52
$Cu(C_2H_5COO)_2.H_2O$ C.S. f		g_z 2·35	$g_{x,y}$ 2·09 –2·10		$D=0{\cdot}38$, $E\sim0$ $S=1$, see text	Ab1 53
$Cu(C_6H_4)_4(CN)_8$ C.S. g	270 to 20	$g_\parallel$ 2·165 $\pm$0·004	$g_\perp$ 2·045 $\pm$0·003		$\psi=+121°$, $\alpha=45°$; in crystal diluted with Zn, $A=0{\cdot}020_8\pm0{\cdot}001$, $B=0{\cdot}003_1\pm0{\cdot}001$	In2 54 Be 55

† Vi 53 show that the results are temperature dependent, with the mean g (powder) at 90°K 1% higher than at 290°K.

a. $Cu[(CH_3CO)_2CH]_2$. Cox and Webster (1935): monoclinic, $M=2$. Each Cu is surrounded by four O in a plane.

b. $CuCl_2.2H_2O$. Harker (1936): orthorhombic, $M=2$. Each Cu is surrounded by a planar array of two H_2O and two Cl. The distorted octahedron is completed by two more Cl, belonging to other Cu; these two Cl are considerably further away.

c. $CuSO_4.5H_2O$. Beevers and Lipson (1934): triclinic, $M=2$. Each Cu is surrounded by a distorted octahedron of four H_2O and two O, with tetragonal symmetry. The two tetragonal axes are nearly at right-angles to each other, and so the crystal is magnetically uniaxial; the susceptibility axis (minimum) makes angles 154°, 64°, 51° with the triclinic axes (a, b, c).

d. $(K, NH_4)_2CuCl_4.2H_2O$. Chrobak (1934), Wyckoff (1951): tetragonal, $M=2$. Each Cu is surrounded by four Cl in a rhomb in the aa-plane, of which two are considerably further from the Cu than the other two, and two H_2O on the normal. The two magnetic complexes have the same y-axis (c), but interchanged z, x axes along [110], [1$\bar{1}$0], defined with respect to Chrobak's unit cell, i.e. z and x are [100] and [010] using the morphological axes of Groth (1906).

e. $Cu(CH_3COO)_2.H_2O$. van Niekerk and Schoening (1953): monoclinic, $M=4$ of the bimolecular units. The four magnetic complexes are equivalent in pairs; description of axes as for C.S. 1. The Cu lie in close pairs (2·6 Å), and each Cu is surrounded by a distorted octahedron of four O, one H_2O and the other Cu.

f. $Cu(C_2H_3COO)_2.H_2O$. Groth (1910): monoclinic. Paramagnetic resonance shows that there are six magnetic complexes (each with two Cu) in unit cell, of which three are derived from the other three by reflection in ac plane. Each complex has approximate tetragonal symmetry, with axes (z) given by $\psi_1=106°, \psi_2=158°, \psi_3=174°$, $a_1=28\frac{1}{2}°, a_2=16°, a_3=52°$; here ψ_i is angle between c and plane containing two z_i axes, and a_i is angle between z and ac plane ($i=1, 2, 3$).

g. $Cu(C_6H_4)_4(CN)_8$. Robertson (1935): monoclinic, $M=2$. Each Cu is surrounded by a square of four N. Description of axes as for C.S. 1.

Paramagnetic resonance absorption has also been observed in the following compounds, unless stated otherwise at room temperature; the g-values are given in brackets: $CuBr_2$ (2·23), $CuC_4H_4O_6.3H_2O$ (1·97), $Cu(C_2H_3O_2)_2.2H_2O$ (2·09), $Cu(C_5H_5N)_4(NO_2)_2.H_2O$ (2·04), $Cu[(CH_3)_2C(CO)_2CH_2C_5H_{11}]_2$ (2·04), $Cu[C_6H_3(CH_3)_2(CO)_2CH_3]_2$ (2·02), $Cu(C_7H_5O_2)_2.2H_2O$ (1·97), all La 51; $Cu(HCOO)_2.2H_2O$, $4H_2O$ (2·32 to 2·06), Ab1 53; CuF_2 (2·15), La 51; CuF_2 basic (2·08), Ra 52; Cu $(NH_3)_4Cl_2.H_2O$, Ba1 48; CuS, Ko 52; $CuSO_4$, Ba1 48; $CuSiF_6.4H_2O$, ($g_\parallel=2·40$, $g_\perp=2·10$), Yo 54; $CuWO_4.2H_2O$ (2·17), La 51; $2CuC_2O_4.H_2O$ (2·07), La 51; $Cu_3Co_2(CN)_{12}$, $\frac{1}{4}T=290$–12, (2·17) Pe2 54; copper chlorophylls, $T=270$, 90, (2·05 ± 0·01), In2 54; cupric di-o-phenanthroline persulphate, $T=290$, 90, 20 ($g_\parallel=2·3_0$, $g_\perp=2·05$), Bo5 53.

Mo^v

Formula	T		$g_\parallel$	$g_\perp$	A	B	Remarks	Ref.
$K_2[InCl_6]2H_2O$ d. 200	290 to 20	I	1·951 ±0·005	1·939 ±0·006	0·0079 ±0·0001	0·00385 ±0·00010	Two magnetically inequivalent complexes, I and II, in unit cell. Hyperfine structure from Cl nuclei also observed	Ow2 u
		II	1·959 ±0·004	1·939 ±0·006	0·0077 ±0·0001	0·00385 ±0·00010		
$K_2[Mo(CN)_8]$ p	290, 90		$\bar{g}=2·005$ ±0·005					Gr3 53

Table 5.4 lists those naturally abundant isotopes of the atoms given in table 5.3 which have a non-zero value for the nuclear spin I. Hyperfine structures have been measured for Mo, Ru and Ir.

Table 5.4

Z	Atom	A	Natural abundance (%)	I
42	Mo	95, 97	$15{\cdot}7$, $9{\cdot}4_5$	$\frac{5}{2}$
44	Ru	99, 101	$12{\cdot}8$, $17{\cdot}0$	$\frac{5}{2}$
47	Ag	107, 109	$51{\cdot}3_5$, $48{\cdot}6_5$	$\frac{1}{2}$
75	Re	185, 187	$37{\cdot}0_7$, $62{\cdot}9_3$	$\frac{5}{2}$
77	Ir	191, 193	$38{\cdot}5$, $61{\cdot}5$	$\frac{3}{2}$

It is of particular interest that in three complexes hyperfine structures have also been observed from the nuclei of the attached ' diamagnetic ' ions. Examples are: $[IrCl_6]^{2-}$, structure from ^{35}Cl, ^{37}Cl; abundances 75·4%, 24·6%; $I=\frac{3}{2}$: and $[IrBr_6]^{2-}$, structure from ^{79}Br, ^{81}Br; 50·5%, 49·5%; $I=\frac{3}{2}$. These structures are direct evidence of the covalent nature of the bonding in complexes of this type, as discussed in § 2.1.

4d^1 *Molybdenum* Mo^{5+}.

The only detailed measurements are for single crystals of $K_3[InCl_6]2H_2O$ containing Mo^{v} as an impurity. The exact form of the paramagnetic complex is not known, although it probably contains oxygen and chlorine atoms, e.g. $[Mo^{v}\ OCl_5]^{2-}$. The resonance spectrum shows that the complex has axial symmetry with nearly ' spin only ' magnetic behaviour.

For the odd isotopes, the lowest levels can be described by the spin Hamiltonian

$$\mathcal{H}=g\,\boldsymbol{\beta}H_zS_z+g_\perp\boldsymbol{\beta}(H_xS_x+H_yS_y)+AS_zI_z+B(S_xI_x+S_yI_y),$$

where $S=\frac{1}{2}$, and $I=\frac{5}{2}$. For the even isotopes, only the term in g is retained. The spectrum also shows signs of a small hyperfine structure from the attached Cl nuclei (cf. Ir^{4+} discussed below), but terms representing this are omitted.

With $H\|z$ and $g_\parallel\boldsymbol{\beta}H\gg B$, the levels are at $\pm\frac{1}{2}g_\parallel\boldsymbol{\beta}H\pm\frac{1}{2}Am$ ($m=\frac{5}{2}, \frac{3}{2}, \ldots, -\frac{5}{2}$). To apply the transformation rules for $H\perp z$, the condition $g_\perp\boldsymbol{\beta}H\gg A,B$ is required.

When $H=0$, the levels are two singlets at $-\frac{1}{4}A\pm\frac{3}{2}B$, and five doublets at $\frac{5}{4}A$, $-\frac{1}{4}A\pm\frac{1}{2}(4A^2+5B^2)^{1/2}$, $-\frac{1}{4}A\pm\frac{1}{2}(A^2+8B^2)^{1/2}$.

With $H\|z$ the susceptibility χ_z is $(3/32)g_z^2/T$.

The value of CT^2/R is 0·252 (35/48) (A^2+2B^2), (A,B in °K), where the factor 0·252 takes account of the fact that only 25·2% of the Mo nuclei have non-zero spin. The tabulated values of A, B are the mean for the two isotopes ^{95}Mo, ^{97}Mo, p. 357.

4d^3 *Molybdenum* Mo^{3+}, 5d^3 *Rhenium* Re^{4+}.

The only detailed measurements have been made on $K_3[MoCl_6]$ mixed with K_3 [In Cl_6] $2H_2O$. The paramagnetic $[MoCl_6]^{3-}$ complex is octahedral with some rhombic distortion, and the ground state is an orbital singlet with $S=\frac{3}{2}$ (cf. 3d^3).

For the odd isotopes the spin Hamiltonian is

$$\mathcal{H}=\boldsymbol{\beta}(g_zH_zS_z+g_xH_xS_x+g_yH_yS_y)+D\{S_z^2-\tfrac{1}{3}S(S+1)\}+E(S_x^2-S_y^2)$$
$$+A_zS_zI_z+A_xS_xI_x+A_yS_yI_y$$

where $S=\frac{3}{2}$, and $I=\frac{5}{2}$. For the even isotopes the term in A is omitted. Neglecting A, when $H=0$ there are two doublets with energies $\pm(D^2+3E^2)^{1/2}$. This zero field splitting is many cm^{-1}, and for fields of a few kilogauss $(D^2+3E^2)^{1/2} \gg g\beta H \gg A$, giving levels when $H \| z$:

$$+(D^2+3E^2)^{1/2} \pm \left\{\frac{2D}{(D^2+3E^2)^{1/2}}+1\right\}(\tfrac{1}{2}g_z\beta H+\tfrac{1}{2}A_z m)$$

$$-(D^2+3E^2)^{1/2} \pm \left\{\frac{2D}{(D^2+3E^2)^{1/2}}-1\right\}(\tfrac{1}{2}g_z\beta H+\tfrac{1}{2}A_z m)$$

where m can take values $m=\frac{5}{2}, \frac{3}{2}, \ldots, -\frac{5}{2}$.

Mo^{iii}

Formula		T		g	D, E	A_z	Remarks	Ref.
$K_3[InCl_6]2H_2O$ d. 200		90, 20	I	1·93 ±0·06	$D \gg 1$ $E/D=0·08$	0·0039 ±0·0005	Many magnetically inequivalent complexes in unit cell—only two, I and II, measured.	Ow2 u
			II	1·93 ±0·06	$D \gg 1$ $E/D=0·15$	0·0039 ±0·0005		
K_2MoCl_5	p	290		1·76				Ra 51
$K_2MoCl_5.H_2O$	p	290 to 20		1·96				Gr3 53
$KMoF_4$	p	14		1·95				Gr3 53

Re^{iv}

$K_2[PtCl_6]$; C.S. 8; d. 200 $T=90, 20$ $g=1·8$ (?)

Six lines with centre of gravity at $g=1·8$: possibly hyperfine structure from ^{185}Re, ^{187}Re. Ref. Gr3 53.

4d^5 *Ruthenium* Ru^{3+}, 5d^5 *Iridium* Ir^{4+}.

The available data are for covalently bound octahedral complexes; there is only a single unpaired electron and the effective spin $S'=\frac{1}{2}$ (see § 2.1).

(*a*) 4d^5 Ru^{3+}.

The most detailed results are for diluted crystals of $[Ru(NH_3)_6]Cl_3$. There are three types of the octahedral $[Ru(NH_3)_6]^{3+}$ complex in this salt, with different rhombic distortions. For each type the spin Hamiltonian for the odd isotopes is

$$\mathcal{H}=\beta(g_zH_zS_z+g_xH_xS_x+g_yH_yS_y)+A_zS_zI_z+A_xS_xI_x+A_yS_yI_y,$$

where $S=\frac{1}{2}$, and $I=\frac{5}{2}$. For the even isotopes only the term in g is retained. With $H \| z$ and $g\beta H \gg A$, the levels are at $\pm\frac{1}{2}g_z\beta H \pm \frac{5}{2}A_z m$, where m can take values $\frac{5}{2}, \frac{3}{2}, \ldots, -\frac{5}{2}$. When $H=0$, the levels are approximately two singlets at $-\frac{1}{4}A \pm \frac{5}{2}B$, and five doublets at $\frac{5}{4}A$, $-\frac{1}{4}A \pm \frac{1}{2}(4A^2+5B^2)^{1/2}$, $-\frac{1}{4}A \pm \frac{1}{2}(A^2+8B^2)^{1/2}$, where $A=A_z$, $B=\frac{1}{2}(A_x+A_y)$.

The value of CT^2/R is 0·298 $(35/48)(A_x^2+A_y^2+A_z^2)$ (A in °K), where the factor 0·298 represents the natural abundance of the odd isotopes; in the table the given values of A are the mean for the two isotopes ^{99}Ru and ^{101}Ru.

Ru^{iii}

Formula	T		g_x	g_y	g_z	A_x	A_y	A_z	Remarks	Ref.
$[Co(NH_3)_6]Cl_3$ d. 200	20	I	2·06 ±0·01	2·02 ±0·01	1·72 ±0·01	0·0048 ±0·0002	0·0048 ±0·0002	0·0049 ±0·0002	†	Gr3 53 Gr2 55
		II	1·80 ±0·01	1·90 ±0·01	2·06 ±0·01	0·0048 ±0·0002	0·0048 ±0·0002	0·0050 ±0·0002		
		III	1·15 ±0·01	1·84 ±0·01	2·66 ±0·01	0·0045 ±0·0002	0·0041 ±0·0002	0·0054 ±0·0002	$^{101}A/^{99}A = 1·09 \pm 0·03$	Gr1 52b
$K_3[InCl_6]2H_2O$ d. 100	20		1·0 ±0·1	1·22 ±0·02	3·24 ±0·02					Ow2 u
$[Ru(NH_3)_6]Cl_3$, $3HgCl_2$	20		2·21	2·05	1·5 ±0·1					Gr3 53

† Monoclinic: 3 pairs of magnetically inequivalent complexes of types I, II and III. Members of each pair have reflection symmetry in *ac* plane.

Paramagnetic resonance absorption has also been observed in undiluted crystals of $[Ru(NH_3)_6]Cl_3$ and $[Ru(NH_3)_5Cl]Cl_2$ at $T=20$, Gr3 53.

Ir^{iv}

Formula	T	g	Ir h.f.s.	Cl or Br h.f.s.	Remarks	Ref.
$K_2[PtBr_6]$ C.S. 8; d. 200	20	$g_{\parallel}$ 1·60 ±0·10			3 inequivalent complexes, with axes ‖ edges of cubic unit cell	Gr1 54
		$g_{\perp}$ 1·87 ±0·04				
$K_2[PtCl_6]$ C.S. 8; d. 200	20	$g_x=g_y=g_z$† =1·78 ±0·02			All complexes equivalent	Gr3 53
$(NH_4)_2[PtCl_6]$ C.S. 8; d. 200	20	$g_x=g_y=g_z$ =1·775 ±0·001	$A_x=A_y=A_z$ =0·00265 ±0·00010	$A_1'=A_2'=A_3'$ =0·00088 ±0·00004	All complexes equivalent; x, y, z ‖ edges of cubic unit cell, and axes of Cl_6 octahedron. A' for ^{35}Cl	Gr1 54
$Na_2[PtBr_6]6H_2O$ C.S. 8; d. 200	20	g_x 2·25 ±0·02	A_x 0·00255 ±0·00010	A_1' 0·0057 ±0·0002	All complexes equivalent; x, y, z ‖ axes of Br_6 octahedron. A' mean for ^{79}Br, ^{81}Br	Gr1 54
		g_y 2·21 ±0·02	A_y 0·00255 ±0·00010	A_2' 0·0057 ±0·0002		
		g_z 0·75 ±0·10				
$Na_2[PtCl_6]6H_2O$ C.S. 8; d. 200	20	g_x 2·20 ±0·02	A_x 0·00255 ±0·00010	A_1' 0·00116 ±0·00004	All complexes equivalent; x, y, z ‖ axes of Cl_6 octahedron. A' for ^{35}Cl	Gr1 54
		g_y 2·07 ±0·02	A_y 0·00255 ±0·00010	A_2' 0·00107 ±0·00004		
		g_z 1·05 ±0·02	A_z 0·0024 ±0·0001	$A_3'<0·0005$		

† This g-value is more accurate than the value given in ref.

(*b*) $5d^5$ Ir^{4+}.

For the compounds in the table above, the spin Hamiltonian is as for Ru^{3+}, except for six additional terms, each representing the splitting from one Cl or Br nucleus. (See § 2.1; for detailed theory Stevens 1953, for a fuller discussion Griffiths and Owen 1954). The most important terms in the complete Hamiltonian are then

$$\mathcal{H}=\beta(g_xH_xS_x+g_yH_yS_y+g_zH_zS_z)+A_xS_xI_x+A_yS_yI_y+A_zS_zI_z$$
$$+A_1'S_x[(I_x)_1+(I_x)_4]+A_2'S_y[(I_y)_2+(I_y)_5]+A_3'S_z[(I_z)_3+(I_z)_6],$$

where $S=\frac{1}{2}$, $I=\frac{3}{2}$ for Ir, and $I_1=I_2=I_3=I_4=I_5=I_6=\frac{3}{2}$ for the Cl or Br nuclei situated on the x, y, z, $-x$, $-y$, $-z$ axes.

With $H\|z$ and $g\beta H\gg A$, A_3', the energy levels are at

$$\pm\tfrac{1}{2}g_z\beta H\pm\tfrac{1}{2}A_zm\pm\tfrac{1}{2}A_3'(m_3+m_6),$$

where m, m_3 and m_6 can each independently take the values $\frac{3}{2}$, $\frac{1}{2}$, $-\frac{1}{2}$, $-\frac{3}{2}$.

For a general direction of H, the levels are more complicated, since all six of the halogen nuclei contribute. We only consider the simple case of a regular octahedral complex for which $g_x=g_y=g_z=g$, $A_x=A_y=A_z=A$, and $A_1'=A_2'=A_3'=A'$. Then, if H has direction cosines (l', m', n') with respect to (x, y, z) and $g\beta H\gg A$, A', the levels are at

$$\pm\tfrac{1}{2}g\beta H\pm\tfrac{1}{2}Am\pm\tfrac{1}{2}A'[l'(m_1+m_4)+m'(m_2+m_5)+n'(m_3+m_6)],$$

where m, m_1, m_2, m_3, m_4, m_5 and m_6 can each independently take the values $\frac{3}{2}$, $\frac{1}{2}$, $-\frac{1}{2}$, $-\frac{3}{2}$.

When $H=0$, the levels are, very approximately, two singlets at $-\frac{1}{4}A\pm B$, and three doublets at $\frac{3}{4}A$ and $-\frac{1}{4}A\pm\frac{1}{2}(A^2+3B^2)^{1/2}$, where $A=A_z$ and $B=\frac{1}{2}(A_x+A_y)$, and the splitting A' from the halogen nuclei has been neglected. These approximate formulae do not apply to $[IrBr_6]^{2-}$, for which $A'>A$. With the same approximation, the value of CT^2/R is $\frac{5}{16}(A_x^2+A_y^2+A_z^2)$ (A in °K). In the table, the given values of A are the mean for the two isotopes.

$4d^9$ *Silver* Ag^{2+}.

Measurements have been made on the powders of two argentic compounds.

		Ag^{II}		
		T	*g*	Ref.
Argentic di-o-phenanthroline persulphate	p	290, 90, 20	$g_\parallel$ 2·18, $g_\perp$ 2·04	Bo5 53
$Ag(C_5H_5N)_4S_2O_8$	p	297, 20	$2{\cdot}08_3$ ($g_\parallel$ 2.18, $g_\perp$ 2·04)	Gi 54
Cd $(C_5H_5N)_4S_2O_8$ d. 20	p	4·2	2·06	Gi 54

5.5. *The* 4f *Group*

For the magnetic complexes of the 4f group, the splittings of the free ion levels by the crystal field are only of order 100 cm^{-1} (see § 2.2). Since at high temperatures rapid spin–lattice relaxation broadens the lines, low temperatures are generally

required to examine the spectrum. (The case of $4f^7$, for which the free ion is in an S-state, forms an exception, and these remarks do not apply.) At such low temperatures (20°K or less) in the crystals investigated only the lowest doublet is populated (except in cerium ethylsulphate, in which a second doublet has been observed), and resonance is observed between its two levels. The effective spin S', defined by equating $2S'+1$ to the number of levels of the ground state, neglecting nuclear interactions, is thus $\frac{1}{2}$. The two levels are degenerate (Kramers doublet) if there is an odd number of electrons; for an even number they are split, but owing to the relatively high (trigonal) symmetry of the complexes investigated, this splitting is small. The resonance results for the 4f group thus fall into three classes, corresponding to the number of 4f electrons: (1) odd number, except $4f^7$, (2) even number, (3) $4f^7$.

Most of the detailed results have been obtained using ethylsulphate and double nitrate crystals. For both of these crystals the magnetic complex has trigonal symmetry; one would not necessarily expect the results for other crystals to be similar. The ions for which resonance has been observed are listed in table 5.5.

Table 5.5

No. of 4f electrons	Ion	Free ion ground state	S'	Remarks
1	Ce^{3+}	$^2F_{5/2}$	$\frac{1}{2}$	one or two Kramers doublets
2	Pr^{3+}	$^3H_4$	$\frac{1}{2}$	split doublet
3	Nd^{3+}	$^4I_{9/2}$	$\frac{1}{2}$	Kramers doublet
5	Sm^{3+}	$^6H_{5/2}$	$\frac{1}{2}$	Kramers doublet
7	Eu^{2+}, Gd^{3+}	$^8S_{7/2}$	$\frac{7}{2}$	
8	Tb^{3+}	$^7F_6$	$\frac{1}{2}$	split doublet
11	Er^{3+}	$^4I_{15/2}$	$\frac{1}{2}$	Kramers doublet

Table 5.6 lists those naturally abundant isotopes of the atoms given in table 5.5 which have a non-zero nuclear spin I. The corresponding hyperfine structure in the resonance spectrum has been observed for each of these isotopes, except those of gadolinium.

Table 5.6

Z	Atom	A	Natural abundance (%)	I
59	Pr	141	100	$\frac{5}{2}$
60	Nd	143, 145	12·2, 8·3	$\frac{7}{2}$
62	Sm	147, 149	$15{\cdot}0_7$, $13{\cdot}8_4$	$\frac{7}{2}$
63	Eu	151, 153	47·77, 52·23	$\frac{5}{2}$
64	Gd	155, 157	14·73, 15·68	—
65	Tb	159	100	$\frac{3}{2}$
68	Er	167	$22{\cdot}9_4$	$\frac{7}{2}$

$4f^1$ *Cerium* Ce^{3+}.

The results (see table below) are interpreted in terms of the following spin Hamiltonian, with $S=\frac{1}{2}$:

$$\mathcal{H}=g_\parallel\beta H_z S_z+g_\perp\beta(H_x S_x+H_y S_y).$$

The energy levels are simply given by $\pm\frac{1}{2}g\beta H$, where $g^2=g_\parallel{}^2\cos^2\theta+g_\perp{}^2\sin^2\theta$ and θ is the angle which H makes with the axis.

In the diluted ethylsulphate there is a second doublet b, 3 ± 1 cm^{-1} above the ground doublet a; doublet b has also been observed by paramagnetic resonance, and the same spin Hamiltonian is used for it. In the undiluted salt, the level order is reversed, b being the ground doublet. Using the double nitrate, only doublet a has been observed.

Ce^{3+}

Formula	T	$g_\parallel{}^a$	$g_\perp{}^a$	$g_\parallel{}^b$	$g_\perp{}^b$	Remarks	Ref.
$Mg_3Ce_2(NO_3)_{12}.24H_2O$ } $Mg_3La_2(NO_3)_{12}.24H_2O$ C.S. 3; d. 20	4·2	0·25 $\pm$0·05	1·84 $\pm$0·02				Co2 53
$Ce(C_2H_5SO_4)_3.9H_2O$ C.S. 9	2·5	1·0 $\pm$0·2	2·25 $\pm$0·2	3·80 $\pm$0·04	$\leqslant$0·4	Doublet b lowest	Bo1 51
$La(C_2H_5SO_4)_3.9H_2O$ C.S. 9; d. 200	4·2	0·955 $\pm$0·005	2·185 $\pm$0·01	3·72 $\pm$0·01	0·20 $\pm$0·05	Doublet b 3 ± 1 cm^{-1} above a	Bo1 51 Bo1 u

Paramagnetic resonance absorption has also been reported for $Ce_2(C_2O_4)_3.9H_2O$ ($T=290$, $g=1{\cdot}95$), La 51, 52; $Ce_2(SO_4)_3$ ($T=290$, 90) and $Ce_2(SO_4)_3.8H_2O$ ($T=90$), Ku3 51.

$4f^2$ *Praseodymium* Pr^{3+}.

Since the number of electrons is even, the lowest doublet need not be degenerate. The ethylsulphate and double nitrate have trigonal symmetry, and it so happens that a trigonal field does not split the doublet. Owing to distortions and thermal fluctuations, however, the symmetry is not truly trigonal, and there is a small random splitting. This is represented by the term Δ in the spin Hamiltonian, which is

$$\mathcal{H}=g_\parallel\beta H_zS_z+AS_zI_z+\Delta_xS_x+\Delta_yS_y \quad (S=\tfrac{1}{2},\ I=\tfrac{5}{2}).$$

Here $g_\perp$ has been taken to be zero, and the quadrupole interaction term $P(I_z{}^2-\frac{35}{12})$, which may be required, has been omitted since no information about it has yet been obtained.

The energy levels are $\pm\frac{1}{2}(\{g_\parallel\beta H\cos\theta+Am\}^2+\Delta^2)^{1/2}$ ($m=\frac{5}{2},\ \frac{3}{2},\ \ldots,\ -\frac{5}{2}$), where θ is the angle which H makes with the axis, and $\Delta^2=\Delta_x{}^2+\Delta_y{}^2$. The value of Δ is not sharply defined; it has a two-dimensional gaussian distribution centred on zero. The value of CT^2/R is

$$\tfrac{1}{4}\Delta^2+\tfrac{35}{48}A^2\ (+\tfrac{56}{9}P^2) \quad (\Delta,\ A,\ P \text{ in } {}^\circ\text{K}).$$

Pr^{3+}

Formula	T	$g_\parallel$	$g_\perp$	Δ	A	Remarks	Ref.
$Mg_3La_2(NO_3)_{12}.24H_2O$ C.S. 3; d. 100	4.2	1·55 $\pm$0·02			0·077 $\pm$0·002		Co1 55
$La(C_2H_5SO_4)_3.9H_2O$ C.S. 9; d. 200	20	1·69 $\pm$0·01	<0·3	~0·04	0·083 $\pm$0·001		Bl13 52
$PrCl_3.7H_2O$	4	3·02	2·23			triclinic $M_m=2$	Da 51

4f³ *Neodymium* Nd^{3+}.

For the crystals investigated (see table below) the spectrum of the odd isotopes is interpreted in terms of the following spin Hamiltonian, with $S=\frac{1}{2}$ and $I=\frac{7}{2}$:

$$\mathcal{H}=g_\parallel\beta H_z S_z+g_\perp\beta(H_x S_x+H_y S_y)+AS_zI_z+B(S_xI_x+S_yI_y).$$

For the even isotopes, only the term in g is retained. With $H\parallel z$ and $g_\parallel\beta H\gg B$, the levels are at

$$\pm\tfrac{1}{2}g_\parallel\beta H\pm\tfrac{1}{2}Am\ (m=\tfrac{7}{2},\ \tfrac{5}{2},\ \ldots,\ -\tfrac{7}{2}).$$

To apply the transformation rules for $H\perp z$, the condition $g_\perp\beta H\gg A,\ B$ is required. In zero field, there are two singlets at $-\frac{1}{4}A\pm 2B$, and seven doublets at $\frac{7}{4}A,\ -\frac{1}{4}A\pm\frac{1}{2}(9A^2+7B^2)^{1/2},\ -\frac{1}{4}A\pm(A^2+3B^2)^{1/2},\ -\frac{1}{4}A\pm\frac{1}{2}(A^2+15B^2)^{1/2}$.

The value of CT^2/R is $\frac{21}{16}\{0{\cdot}122(^{143}A^2+2^{143}B^2)+0{\cdot}083(^{145}A^2+2\ ^{145}B^2)\}$ (A,B in °K), where the factors 0·122 and 0·083 have been introduced to take account of the natural abundances.

4f⁵ *Samarium* Sm^{3+}.

The spin Hamiltonian, etc., is the same as for Nd^{3+}. The value of CT^2/R is $\frac{21}{16}\{0{\cdot}151(^{147}A^2+2^{147}B^2)+0{\cdot}138(^{149}A^2+2^{149}B^2)\}$ (A, B in °K), where the factors 0·151 and 0·138 have been introduced to take account of the natural abundances (see table below).

4f⁷ *Gadolinium* Gd^{3+}, *Europium* Eu^{2+}.

(*a*) Gd^{3+}.

A full analysis of the spectrum has been carried out for three crystals (see table below). The hyperfine structure due to the odd isotopes has not been observed, and the nuclear interactions are therefore omitted from the spin Hamiltonian. Owing to the large value of $S\ (\frac{7}{2})$, a number of new terms have to be introduced into the spin Hamiltonian, and this is now most conveniently written in the form:

$$\mathcal{H}=g\beta(H_zS_z+H_xS_x+H_yS_y)+B_2^0P_2^0+B_2^2P_2^2+B_4^0P_4^0+B_6^0P_6^0+B_6^6P_6^6$$

where each P_n^m is an operator; for a discussion of the Hamiltonian and details of the operators see Elliott and Stevens (1953). The operator P_2^2, which is not given explicitly by them, is $(S_x^2-S_y^2)$. The coefficients B_n^m are determined by experiment; for numerical convenience, the following nomenclature is used:

$$b_2^0=3B_2^0,\ b_2^2=3B_2^2,\ b_4^0=60B_4^0,\ b_6^0=1260B_6^0,\ b_6^6=1260B_6^6.$$

It may be remarked that b_2^0 and $\frac{1}{3}b_2^2$ are respectively the parameters D and E which have been used whenever $S'>\frac{1}{2}$, and that $3b_4^0=F$, the parameter which has been used for $S'=\frac{5}{2}$ (see § 5.3).

For the ethylsulphate and double nitrate, b_2^2 is zero. For these two crystals with $H\parallel z$ and neglecting the tiny effect of b_6^6, the levels are at

$$\left.\begin{aligned}&\pm\tfrac{7}{2}g\beta H+7b_2^0+7b_4^0+b_6^0\\&\pm\tfrac{5}{2}g\beta H+b_2^0-13b_4^0-5b_6^0\\&\pm\tfrac{3}{2}g\beta H-3b_2^0-3b_4^0+9b_6^0\\&\pm\tfrac{1}{2}g\beta H-5b_2^0+9b_4^0-5b_6^0\end{aligned}\right\}\text{(A)}$$

With $H\perp z$, the further conditions $g\beta H\gg b_2^0,\ b_4^0,\ b_6^0$ are required, and b_2^0 is replaced by $-\frac{1}{2}b_2^0$, b_4^0 by $\frac{3}{8}b_4^0$, and b_6^0 by $\frac{1}{16}(b_6^6\cos 6\phi-5b_6^0)$ where ϕ is the angle between

Nd^{3+}

Formula	T	$g_\parallel$	$g_\perp$	Isotope	A	B	Remarks	Ref.
$Mg_3La_2(NO_3)_{12}.24H_2O$ C.S. 3; d. 100	4·2	0·45 ±0·05	2·72 ±0·02	^{143}Nd	0·0052 ±0·0005	0·0312 ±0·0001		Co1 55
				^{145}Nd	0·0032 ±0·0003	0·0194 ±0·0001		
$La(C_2H_5SO_4)_3.9H_2O$ C.S. 9; d. 200	20	3·535 ±0·001	2·072 ±0·001	^{143}Nd	0·03803 ±0·00001	0·01989 ±0·00005	$^{143}A/^{145}A = 1·6083 \pm 0·0012$	Bl14 54
				^{145}Nd	0·02364 ±0·00001	0·01237 ±0·00005		

Paramagnetic resonance absorption has also been reported for $Nd_2(SO_4)_3$ and $Nd_2(SO_4)_3.8H_2O$ ($T = 290, 90$), Ku3 51; Nd_2O_3 ($T = 290$, $g = 3·2$), Ga 52a.

Sm^{3+}

Formula	T	$g_\parallel$	$g_\perp$	Isotope	A	B	Remarks	Ref.
$Mg_3Sm_2(NO_3)_{12}.24H_2O$ C.S. 3	4·2	0·76 ±0·01	0·40 ±0·05	^{147}Sm	0·0346 ±0·0005	<0·010		Co1 55
				^{149}Sm	0·0287 ±0·0005	<0·010		
$La(C_2H_5SO_4)_3.9H_2O$ C.S. 9; d. 100	4·2	0·596 ±0·002	0·604 ±0·002	^{147}Sm	0·0060 ±0·0001	0·0251 ±0·0001	$P < 0·0004$	Bo4 52
				^{149}Sm	0·0049 ±0·0001	0·0205 ±0·0001	$^{147}A/^{149}A = 1·222 \pm 0·008$	

Gd^{3+}

Formula	T	g	b_2^0	b_2^2	b_4^0	b_6^0	b_6^6	Ref.
$Mg_3Bi_2(NO_3)_{12}.24H_2O$ C.S. 3; d. 100	90, 20	1·992 ±0·003	+0·0124 ±0·0001		+0·00009 ±0·00001	+0·00006 ±0·00001	0·0012 ±0·0001	Tr 53
$La(C_2H_5SO_4)_3.9H_2O$ C.S. 9; d. 200	90	1·990 ±0·002	$+0·0204_7$ ±0·0002		$-0·00039_6$ ±0·00003	$+0·00006_2$ ±0·00001	+0·00035 ±0·00005	Bl14 54
	20		$+0·0199_8$ ±0·0001		−0·000391 ±0·000015	+0·000053 ±0·000005	+0·00040 ±0·00005	
$Sm_2(SO_4)_3.8H_2O$ C.S. a; d. 200	300		†(+)0·0633 ±0·0005	(+)0·038 ±0·005	(−)0·0013 ±0·0003			Bo3 54

† We are informed that the value of b_2^0 quoted in Bo3 54 is incorrect.

a. $Sm_2(SO_4)_3.8H_2O$. Zachariasen (1935): monoclinic, $M=4$, i.e. 8 Sm in unit cell. With Gd, the magnetic complexes are equivalent in fours. The axes are given by $\psi_z=28°$, $a_z=\pm35°$, $\psi_y=0°$, $a_y=\pm52°$, where ψ_i is angle between c and plane containing i axes, and a_i is angle between i axis and ac plane ($i=z, y$).

Paramagnetic resonance has also been observed with $La(NO_3)_3.6H_2O$ (d. 10), Ga 52b; $GdCl_3.6H_2O$, La 51; $Gd(BrO_3)_3.9H_2O$, Bl6 51.

H and Ox, Ox being the normal to a prism face of the crystal. In zero field, if $b_2^0 \gg b_6^6$ (a good approximation), the levels are as given in formulae (A) above, with $H=0$.

For the sulphate, with $H\|z$ and $g\beta H \gg b_2^2$, the levels are as given in formulae (A) above, with $b_6^0=0$. With $H\|x, y$, the further conditions $g\beta H \gg b_2^0, b_4^0$ are required; for $H\|x$, b_2^0 is replaced by $-\frac{1}{2}(b_2^0-b_2^2)$; for $H\|y$, b_2^0 is replaced by $-\frac{1}{2}(b_2^0+b_2^2)$. With $H=0$, if $b_2^0 \gg b_2^2$ (a very rough approximation), there are four doublets at $7b_2^0+7b_4^0+\frac{7}{30}X$, $b_2^0-13b_4^0+\frac{5}{6}X$, $-3b_2^0-3b_4^0+\frac{31}{10}X$, $-5b_2^0+9b_4^0-\frac{25}{6}X$, where $X=(b_2^2)^2/b_2^0$.

For all three crystals, with $H\|z$ the susceptibility χ_z is

$$\frac{63}{32}\frac{g^2}{T}\left(1-\frac{4b_2^0}{T}\right) \quad (b_2^0 \text{ in } {}^\circ\text{K}).$$

The value of CT^2/R is $21(b_2^0)^2+7(b_2^2)^2+77(b_4^0)^2+33(b_6^0)^2+\frac{2}{7}(b_6^6)^2$ (all parameters in °K).

(b) Eu^{2+}.

The powder spectrum has been observed of Eu^{2+} in SrS. The surroundings of the europium ion have cubic symmetry; the results indicate that g and A are isotropic, but the full spin Hamiltonian has not been given.

Eu^{2+}

	T	g	^{151}A	^{153}A	$^{151}A/^{153}A$	Ref.
SrS p	290	$1{\cdot}991_2$	0·00308	0·00138	$2{\cdot}23_5$	Bl10 55
d. ~300		±0·001	±0·00002	±0·00002	±0·04	

$4f^8$ *Terbium* Tb^{3+}.

Since the number of electrons is even, the lowest doublet need not be degenerate. Only one crystal has been examined (ethylsulphate), and the trigonal field in this crystal slightly splits the doublet. The results are interpreted in terms of the following spin Hamiltonian, where $g_\perp$ has been taken to be zero, with $S=\frac{1}{2}$ and $I=\frac{3}{2}$:

$$\mathcal{H}=g_\| \beta H_z S_z + A S_z I_z + \Delta S_x.$$

The quadrupole interaction term $P(I_z^2-\frac{5}{4})$, which may be required, has been omitted since no information about it has yet been obtained.

The energy levels are at $\pm\frac{1}{2}(\{g_\|\beta H \cos\theta + Am\}^2+\Delta^2)^{1/2}$ where θ is the angle which H makes with the axis, and m takes the values $\frac{3}{2}$, $\frac{1}{2}$, $-\frac{1}{2}$, $-\frac{3}{2}$.

The value of CT^2/R is $\frac{5}{16}A^2+\frac{1}{4}\Delta^2$ $(+P^2)$ $(A, \Delta, P$ in °K).

Tb^{3+}

	T	$g_\|$	$g_\perp$	Δ	A	Ref.
$Y(C_2H_5SO_4)_3.9H_2O$	20	17·72	<0·3	0·387	0·209	Ba4 55
C.S. 9; d. ~1000		±0·02		±0·001	±0·002	

$4f^{11}$ *Erbium* Er^{3+}.

Only the ethylsulphate has been examined. For the odd isotope, the results are interpreted in terms of the following spin Hamiltonian, with $S=\frac{1}{2}$ and $I=\frac{7}{2}$:

$$\mathcal{H}=g_\|\beta H_z S_z + g_\perp \beta(H_x S_x + H_y S_y) + A S_z I_z + B(S_x I_x + S_y I_y) + P(I_z^2 - 21/4).$$

For the even isotopes, only the term in g is retained. With $H\|z$ and $g_\|\beta H \gg B$, the levels are at $\pm\frac{1}{2}g_\|\beta H \pm \frac{1}{2}Am + P(m^2-21/4)$, $(m=\frac{7}{2}, \frac{5}{2}, \ldots, -\frac{7}{2})$. To apply the transformation rules for $H\perp z$, the conditions $g_\perp\beta H \gg A$, B and $B \gg P$ are required. In zero field, the levels are two singlets at $-\frac{1}{4}A-5P\pm 2B$, and seven doublets at $\frac{7}{4}A+7P$, $-\frac{1}{4}A+4P\pm\frac{1}{2}(9\{A-2P\}^2+7B^2)^{1/2}$, $-\frac{1}{4}A-P\pm(\{A-2P\}^2+3B^2)^{1/2}$, $-\frac{1}{4}A-4P\pm\frac{1}{2}(\{A-2P\}^2+15B^2)^{1/2}$.

The value of CT^2/R is $0{\cdot}229[\frac{21}{16}(A^2+2B^2)+21P^2]$ (A, B, P in °K) where the factor 0·229 has been introduced to take account of the natural abundance.

Er^{3+}

	T	$g_\|$	$g_\perp$	A	B	P	Ref.
$La(C_2H_5SO_4)_3.9H_2O$	4	1·47	$8{\cdot}8_5$	0·0052	0·0314	0·0030	Bo2 52
C.S. 9; d. 200		±0·03	±0·2	±0·0001	±0·0001	±0·0003	

5.6. *The* 5f–(6d) *Group*

The investigation by paramagnetic resonance of the 5f group is still in its infancy, and there are very few detailed paramagnetic resonance results. These are on the neptunyl $(NpO_2)^{2+}$ and plutonyl $(PuO_2)^{2+}$ ions, having respectively one and two 5f electrons; the uranyl $(UO_2)^{2+}$ ion has no 5f electrons, and is diamagnetic. The theory of the paramagnetism of these ions is discussed by Eisenstein and Pryce (1955). Powder measurements on some uranium halides have also been reported.

$5f^1$ *Neptunyl* $(NpO_2)^{2+}$.

Only one compound, a double nitrate, has been investigated in detail. The spectrum of the odd isotope ^{237}Np is interpreted in terms of the following spin Hamiltonian, with $S=\frac{1}{2}$ and $I=\frac{5}{2}$:

$$\mathcal{H}=g_\|\beta H_z S_z+g_\perp\beta(H_x S_x+H_y S_y)+AS_zI_z+B(S_xI_x+S_yI_y)+P(I_z^2-35/12).$$

With $H\|z$ and $g_\|\beta H \gg B$, the levels are at

$$\pm\tfrac{1}{2}g_\|\beta H \pm \tfrac{1}{2}Am + P(m^2-35/12),\ (m=\tfrac{5}{2}, \tfrac{3}{2}, \ldots, -\tfrac{5}{2}).$$

The transformation rules for $H\perp z$ cannot be applied, as they would require the conditions $g_\perp\beta H \gg A$, B and $B \gg P$. In zero field, there are two singlets at $-\frac{1}{4}A-\frac{5}{3}P\pm\frac{3}{2}B$, and five doublets at

$$\tfrac{5}{4}A+\tfrac{10}{3}P,\ -\tfrac{1}{4}A+\tfrac{4}{3}P\pm\tfrac{1}{2}(4\{A-2P\}^2+5B^2)^{1/2},\ -\tfrac{1}{4}A-\tfrac{5}{3}P\pm\tfrac{1}{2}(\{A-2P\}^2+8B^2)^{1/2}.$$

The value of CT^2/R is $\frac{35}{48}(A^2+2B^2)+\frac{56}{9}P^2$ (A, B, P in °K).

$(NpO_2)^{2+}$

	T	$g_\|$	$g_\perp$	A	B	P
$(UO_2)Rb(NO_3)_3$	20 to 12	3·40	$0{\cdot}20_5$	$(+)0{\cdot}1654_5$	$0{\cdot}0178_2$	$(-)0{\cdot}0301_5$
C.S. 10		±0·01	±0·02	±0·0002	±0·0002	±0·0003

Remarks	Ref.
Relative sign of A and P directly determined	Bl9 54a
These values are more accurate than those given in the reference.	

$5f^2$ *Plutonyl* $(PuO_2)^{2+}$.

In the two plutonyl compounds which have been examined, the position is similar to that for the trigonal crystals of praseodymium Pr^{3+}. A doublet lies lowest, but since the number of electrons is even, the doublet need not be degenerate. In the crystals investigated (see table below), owing to the symmetry of the magnetic complex the doublet is only split through the effects of crystal distortions and thermal fluctuations. The small random splitting is represented by the term Δ in the spin Hamiltonian, which is

$$\mathcal{H}=g_{\parallel}\beta H_z S_z+AS_zI_z+\Delta_x S_x+\Delta_y S_y,$$

with $S=\frac{1}{2}$, and $I=\frac{1}{2}$ for ^{239}Pu, $I=\frac{5}{2}$ for ^{241}Pu. Here $g_{\perp}$ has been taken to be zero, and the quadrupole interaction term $P(I_z^2-35/12)$, which may be required for ^{241}Pu, has been omitted, since no information about it has yet been obtained. The energy levels are

$$\pm\tfrac{1}{2}(\{g_{\parallel}\beta H\cos\theta+Am\}^2+\Delta^2)^{1/2}\ (m=\tfrac{1}{2},\ -\tfrac{1}{2}\text{ for }^{239}\text{Pu};\ m=\tfrac{5}{2},\ \tfrac{3}{2},\ \ldots,\ -\tfrac{5}{2}\text{ for }^{241}\text{Pu}),$$

where θ is the angle which H makes with the axis, and $\Delta^2=\Delta_x^2+\Delta_y^2$. The value of Δ is not sharply defined; it has a two-dimensional gaussian distribution centred on zero.

The value of CT^2/R is $\frac{1}{16}A^2+\frac{1}{4}\Delta^2$ for ^{239}Pu, and $\frac{35}{48}A^2+\frac{1}{4}\Delta^2$ $(+\frac{5}{9}P^2)$ for ^{241}Pu (A, Δ, P in °K).

$(PuO_2)^{2+}$

Formula	T	$g_{\parallel}$	$g_{\perp}$	^{239}A	^{241}A	Ref.
$(UO_2)Rb(NO_3)_3$	20 to 12	5·32	$\leqslant$0·4	0·0862	0·0609	† Bl9 54b
C.S. 10; d. 17 to 200		±0·02		±0·0005	±0·0004	
$Na(PuO_2)(CH_3COO)_3$	4	$5{\cdot}9_2$	~0			Hu2 54
C.S. a						

† $^{241}A/^{239}A=0{\cdot}706\pm0{\cdot}004$; Δ stated to be small

a $Na(UO_2)(CH_3COO)_3$. Fankuchen (1935): cubic, $M=4$. The linear O–U–O groups lie on the body diagonals of the unit cube; around each U there are also six O from acetate groups, the eight O being arranged in a distorted cube.

Uranium halides.

The following powder measurements have been reported:

	T	g	Ref.
UF_3 p	290, 90	2·4 ($g_{\parallel}=2{\cdot}8-2{\cdot}9$, $g_{\perp}=2{\cdot}1-2{\cdot}2$)	Gh 54
UF_4 p	290	$2{\cdot}1_5$	

ACKNOWLEDGMENTS

This report has been written at the suggestion of Dr. Bleaney; we are indebted to him and Drs. Bagguley and Griffiths for much helpful advice and discussion, and to Mr. H. M. Powell for his comments on the crystallography section. We also wish to acknowledge the assistance given in the compilation of the resonance data by Dr. Krüger and Mr. Meyer-Berkhout, of Heidelberg University†, who

† *Note added in proof.* A compilation by these authors has recently been published in Landolt–Börnstein, *Zahlenwerte und Funktionen*, 6th Edn, Vol. 1, Part IV (Kristalle), p. 956.

kindly made available to us an earlier compilation, and by the various laboratories who informed us of their most recent results. This report has been prepared during the tenure of a Research Lectureship at Christ Church (K.D.B.) and a Fellowship at Magdalen College (J.O.).

References

(a) Crystallography

ASTBURY, W. T., 1926, *Proc. Roy. Soc.* A, **112,** 448.
BARKHATOV, V., 1942, *Acta Phys.-chim. U.R.S.S.*, **16,** 123.
BARKHATOV, V., and ZHDANOV, H., 1942, *Acta Phys.-chim. U.R.S.S.*, **16,** 43.
BEEVERS, C. A., and LIPSON, H., 1934, *Proc. Roy. Soc.* A, **146,** 570.
BEEVERS, C. A., and SCHWARTZ, C. M., 1935, *Z. Kristallogr.*, **91,** 157.
BOKII, G. B., and USIKOV, P. I., 1940, *C. R. Acad. Sci., U.R.S.S.*, **26,** 782.
CHROBAK, L., 1934, *Z. Kristallogr.*, **88,** 35.
COX, E. G., and WEBSTER, K. C., 1935, *J. Chem. Soc.*, 731.
FANKUCHEN, I., 1935, *Z. Kristallogr.*, **91,** 473.
FERRARI, A., 1926, *R. C. Accad. Lincei*, **3,** 224.
GOTTFRIED, C., and NAGELSCHMIDT, J. G., 1930, *Z. Kristallogr.*, **73,** 357.
GROTH, P., 1906, *Chemische Kristallographie* (Leipzig: Engelmann), Vol. 1; 1908, *Ibid.*, Vol. 2; 1910, *Ibid.*, Vol. 3.
HARKER, D., 1936, *Z. Kristallogr.*, **93,** 136.
HOARD, J. L., and STROUPE, J. D., quoted by DIEKE, G. H., and DUNCAN, A. B. F., 1949, *Spectroscopic Properties of Uranium Compounds*, National Nuclear Energy Series, III, 2 (New York: McGraw-Hill).
HOFMANN, W., 1931, *Z. Kristallogr.*, **78,** 279.
KETELAAR, J. A. A., 1937, *Physica*, **4,** 619.
LIPSON, H., 1935, *Proc. Roy. Soc.* A, **151,** 347.
LIPSON, H., and BEEVERS, C. A., 1935, *Proc. Roy. Soc.* A, **148,** 664.
VAN NIEKERK, J. N., and SCHOENING, F. R. L., 1953, *Acta Cryst., Camb.*, **6,** 227.
PAULING, L., 1930, *Z. Kristallogr.*, **72,** 482.
POSPELOV, V. A., and ZHDANOV, G. S., 1947, *J. Phys. Chem., Moscow*, **21,** 405.
POWELL, H. M., and WELLS, A. F., 1935, *J. Chem. Soc.*, 359.
ROBERTSON, J. M., 1935, *J. Chem. Soc.*, 615.
TUTTON, A. E. H., 1916, *Phil. Trans. Roy. Soc.* A, **216,** 1.
WYCKOFF, R. W., 1920, *Amer. J. Sci.*, **50,** 317; 1951, *Crystal Structures* (New York: Interscience Publishers, Inc.), Vol. 2, Chap. 10.
YU, S. H., and BEEVERS, C. A., 1936, *Z. Kristallogr.*, **95,** 426.
ZACHARIASEN, W. H., 1935, *J. Chem. Phys.*, **3,** 197.

(b) Published Paramagnetic Resonance Data

Ab1 ABE, H., 1953, *Phys. Rev.*, **92,** 1572.
Ab2 ABE, H., ONO, K., HAYASHI, I., SHIMADA, J., and IWANAGA, K., 1954, *J. Phys. Soc. Japan*, **9,** 814.
Ab3 ABE, H., and SHIMADA, J., 1953 a, *National Science Report*, Ochanomizu University, Vol. 4, no. 1, 77; 1953 b, *Phys. Rev.*, **90,** 316.
ABRAGAM, A., and PRYCE, M. H. L., 1951, *Proc. Roy. Soc.* A, **205,** 135.
Ar ARAKAWA, T., 1954, *J. Phys. Soc. Japan*, **9,** 790.
Ba1 BAGGULEY, D. M. S., BLEANEY, B., GRIFFITHS, J. H. E., PENROSE, R. P., and PLUMPTON, B. I., 1948, *Proc. Phys. Soc.*, **61,** 551.
Ba2 BAGGULEY, D. M. S., and GRIFFITHS, J. H. E., 1950 a, *Proc. Roy. Soc.* A, **204,** 188; 1950 b, *Ibid.*, **201,** 366; 1952, *Proc. Phys. Soc.* A, **65,** 594.
BAINBRIDGE, K. T., and NIER, A. O., 1954, *Preliminary Report No.* 9, *Nuclear Science Series*, N. R. C. Washington, D.C.
Ba4 BAKER, J. M., and BLEANEY, B., 1952, *Proc. Phys. Soc.* A, **65,** 952; 1955, *Ibid.*, **68,** 257.
Ba5 BAKER, J. M., BLEANEY, B., and BOWERS, K. D., 1955, in the press.
Be BENNETT, J. E., and INGRAM, D. J. E., 1955, *Nature, Lond.*, **175,** 130.
Bi BIJL, D., and ROSE-INNES, A. C., 1953, *Proc. Phys. Soc.* A, **66,** 954.

Bl1 Bleaney, B., 1950 a, *Phys. Rev.*, **78,** 214; 1950 b, *Proc. Roy. Soc.* A, **204,** 203; 1951, *Phil. Mag.*, **42,** 441.

Bl2 Bleaney, B., Bogle, G. S., Cooke, A. H., Duffus, R. J., O'Brien, M. C. M., and Stevens, K. W. H., 1955, *Proc. Phys. Soc.* A, **68,** 57.

Bl3 Bleaney, B., and Bowers, K. D., 1951, *Proc. Phys. Soc.* A, **64,** 1135; 1952, *Proc. Roy. Soc.* A, **214,** 451.

B14 Bleaney, B., Bowers, K. D., and Ingram, D. J. E., 1955, *Proc. Roy. Soc.* A, **228,** 147.

B15 Bleaney, B., Bowers, K. D., and Trenam, R. S., 1955, *Proc. Roy. Soc.* A, **228,** 157.

Bl6 Bleaney, B., Elliott, R. J., Scovil, H. E. D., and Trenam, R. S., 1951, *Phil. Mag.*, **42,** 1062.

B17 Bleaney, B., and Ingram, D. J. E., 1951 a, *Proc. Roy. Soc.* A, **205,** 336; 1951 b, *Ibid.*, **208,** 143; 1952, *Proc. Phys. Soc.* A, **65,** 953.

Bl8 Bleaney, B., Ingram, D. J. E., and Scovil, H. E. D., 1951, *Proc. Phys. Soc.* A, **64,** 601.

Bl9 Bleaney, B., Llewellyn, P. M., Pryce, M. H. L., and Hall, G. R., 1954 a, *Phil Mag.*, **45,** 992; 1954 b, *Ibid.*, **45,** 991.

Bl10 Bleaney, B., and Low, W., 1955, *Proc. Phys. Soc.* A, **68,** 55.

Bl11 Bleaney, B., and Penrose, R. P., 1948, *Proc. Phys. Soc.*, **60,** 395.

Bl12 Bleaney, B., Penrose, R. P., and Plumpton, B. I., 1949, *Proc. Roy. Soc.* A, **198,** 406.

Bl13 Bleaney, B., and Scovil, H. E. D., 1952, *Phil. Mag.*, **43,** 999.

Bl14 Bleaney, B., Scovil, H. E. D., and Trenam, R. S., 1954, *Proc. Roy. Soc.* A, **223,** 15.

I Bleaney, B., and Stevens, K. W. H., 1953, *Rep. Progr. Phys.*, **16,** 108 (London: Physical Society).

Bl15 Bleaney, B., and Trenam, R. S., 1954, *Proc. Roy. Soc.* A, **223,** 1.

Bo1 Bogle, G. S., Cooke, A. H., and Whitley, S., 1951, *Proc. Phys. Soc.* A, **64,** 931.

Bo2 Bogle, G. S., Duffus, H. J., and Scovil, H. E. D., 1952, *Proc. Phys. Soc.* A, **65,** 760.

Bo3 Bogle, G. S., and Heine, V., 1954, *Proc. Phys. Soc.* A, **67,** 734.

Bo4 Bogle, G. S., and Scovil, H. E. D., 1952, *Proc. Phys. Soc.* A, **65,** 368.

Bo5 Bowers, K. D., 1952, *Proc. Phys. Soc.* A, **65,** 860; 1953, *Ibid.*, **66,** 666.

Br Brovetto, P., Cini, G., and Ferroni, S., 1953, *Nuovo Cim.*, **10,** 1325.

Co1 Cooke, A. H., and Duffus, H. J., 1955, *Proc. Roy. Soc.* A, **229,** 407.

Co2 Cooke, A. H., Duffus, H. J., and Wolf, W. P., 1953, *Phil. Mag.*, **44,** 623.

Da Davis, C. F., Kip, A. F., and Malvano, R., 1951, *R. C. Accad. Lincei*, **11,** 77.
Eisenstein, J. C., and Pryce, M. H. L., 1955, *Proc. Roy. Soc.* A, **229,** 20.
Elliott, R. J., and Stevens, K. W. H., 1953, *Proc. Roy. Soc.* A, **219,** 387.

Ga Garafianov, N. S., 1952 a, *Thesis*, Kazan University, quoted by Kozyrev, B. M., 1954, *J. Chim. Phys.*, **51,** 104; 1952 b, *Doklady Akad. Nauk.*, **84,** 923; see *Sci. Abstr.*, 1953, 5569.

Gh Ghosh, S. N., Gordy, W., and Hill, D. G., 1954, *Phys. Rev.*, **96,** 36.

Gi Gijsman, H. M., Gerritsen, H. J., and van den Handel, J., 1954, *Physica*, **20,** 15.

Gr1 Griffiths, J. H. E., and Owen, J., 1952 a, *Proc. Roy. Soc.* A, **213,** 459; 1952 b, *Proc. Phys. Soc.* A, **65,** 951; 1954, *Proc. Roy. Soc.* A, **226,** 96.

Gr2 Griffiths, J. H. E., O'Brien, M. C. M., Owen, J., and Ward, I. M., 1955, to be published.

Gr3 Griffiths, J. H. E., Owen, J., and Ward, I. M., 1953, *Proc. Roy. Soc.* A, **219,** 526.

Ha1 Halliday, D., and Wheatley, J., 1948, *Phys. Rev.*, **74,** 1712.

Ha2 Hayashi, I., and Ono, K., 1953, *J. Phys. Soc. Japan*, **8,** 270.

He Herschberger, W. D., and Leifer, H. N., 1952, *Phys. Rev.*, **88,** 714.

Ho Holden, A. N., Kittel, C., and Yager, W. A., 1949, *Phys. Rev.*, **75,** 1443.

Hu1 Hurd, F. K., Sachs, M., and Herschberger, W. D., 1954, *Phys. Rev.*, **93,** 373.

Hu2 Hutchinson, C. A., and Lewis, W. B., 1954, *Phys. Rev.*, **95,** 1096.

Hu3 Hutchinson, C. A., and Singer, L. S., 1953, *Phys. Rev.*, **89,** 256.

In1 Ingram, D. J. E., 1953 a, *Proc. Phys. Soc.* A, **66,** 412; 1953 b, *Phys. Rev.*, **90,** 711.

In2 Ingram, D. J. E., and Bennett, J. E., 1954, *J. Chem. Phys.*, **22,** 1136.

It Itoh, J., Fujimoto, M., and Ibamoto, H., 1951, *Phys. Rev.*, **83,** 852.
Judd, B. R., 1955, *Proc. Roy. Soc.* A, in the press.

Ki1 Kikuchi, C., Sirvetz, H. M., and Cohen, V. W., 1953, *Phys. Rev.*, **92,** 109.

Ki2 Kip, A. F., Davis, C. F., Jennings, L., Reiner, D., and Malvano, R., 1951, *Nuovo Cim.*, **8,** 683.

Ko Kozyrev, B. M., Salikhov, S. G., and Shamonin, Yu. Ya., 1952, *J. Exp. Theor. Phys.*, **22,** 56, see *Sci. Abst.*, 1953, 3431.

Ku1 Kumagai, H., Ono, K., Hayashi, I., Abe, H., Shimada, J., Shono, H., and Ibamoto, H., 1951, *Phys. Rev.*, **83,** 1077.

Ku2 Kumagai, H., Ono, K., Hayashi, I., and Kambe, K., 1952, *Phys. Rev.*, **87,** 374.

Ku3 Kurenev, V. Ya. and Salikhov, S. G., 1951, *J. Exp. Theor. Phys.*, **21,** 864, see *Sci. Abstr.*, 1952, 5194.

La LANCASTER, F. W., and GORDY, W., 1951, *J. Chem. Phys.*, **19,** 1181; 1952, *Ibid.*, **20,** 740.
Me MEIJER, P. H. E., 1951, *Physica*, **17,** 899.
Ok OKAMURA, T., and DATE, M., 1954, *Phys. Rev.*, **94,** 314.
On1 ONO, K., 1953, *J. Phys. Soc. Japan*, **8,** 802.
On2 ONO, K., KOIDE, S., SEKIYAMA, H., and ABE, H., 1954, *Phys. Rev.*, **96,** 38.
Ow OWEN, J., 1955 a, *Proc. Roy. Soc.* A, **227,** 183; 1955 b, to be published.
Pe1 PENROSE, R. P., and STEVENS, K. W. H., 1950, *Proc. Phys. Soc.* A, **63,** 29.
Pe2 PERAKIS, N., WUCHER, J., and GIJSMAN, H. M., 1954, *C. R. Acad. Sci., Paris*, **239,** 243.
PRYCE, M. H. L., 1950, *Proc. Phys. Soc.* A, **63,** 25.
Ra RAMASESHAN, S., and SURYAN, G., 1951, *Phys. Rev.*, **84,** 593; 1952, *Proc. Indian Acad. Sci., Bangalore*, A, **36,** 211.
Sc SCHNEIDER, E. E., and ENGLAND, T. S., 1951, *Physica*, **17,** 221.
STEVENS, K. W. H., 1952, *Proc. Phys. Soc.* A, **65,** 209; 1953, *Proc. Roy. Soc.* A, **219,** 542.
Si SINGER, L. S., 1955, *J. Chem. Phys.*, **23,** 379.
Ti1 TING, Y., and WILLIAMS, D., 1951, *Phys. Rev.*, **82,** 507.
Ti2 TINKHAM, M., 1955, *Proc. Phys. Soc.* A, **68,** 258.
Tr TRENAM, R. S., 1953, *Proc. Phys. Soc.* A, **66,** 118.
Ub1 UBBINK, J., POULIS, J. A., GERRITSEN, H. J., and GORTER, C. J., 1952, *Physica*, **18,** 361.
Ub2 UBBINK, J., POULIS, J. A., and GORTER, C. J., 1951, *Physica*, **17,** 213.
Vi VITTORELLI, M. B. P., PALMA, M. U., PALEMBO, D., and SANTANGELO, M., 1953, *Ric. Sci.*, **23,** 1423.
VAN VLECK, J. H., 1932 a, *The Theory of Electric and Magnetic Susceptibilities* (Oxford: University Press); 1932 b, *Phys. Rev.*, **41,** 208; 1935, *J. Chem. Phys.*, **3,** 807; 1939, *Ibid.*, **7,** 61.
Wh WHITMER, C. A., WEIDNER, R. T., HSIANG, J. S., and WEISS, P. R., 1948, *Phys. Rev.*, **74,** 1478.
Ya YAGER, W. A., MERRITT, F. R., HOLDEN, A. N., and KITTEL, C., 1949, *Phys. Rev.*, **75,** 1630.
Yo YOKOZAWA, Y., 1954, *Monogr. Res. Inst. Appl. Elect. Hokkaido Univ.*, **4,** 95.
ZAVOISKY, E., 1945, *J. Phys., U.S.S.R.*, **9,** 211.

(c) Unpublished Paramagnetic Resonance Data

Ba1 u BAGGULEY, D. M. S., GRIFFITHS, J. H. E., and OWEN, J.
Ba2 u BAGGULEY, D. M. S., OWEN, J., and WARD, I. M.
Ba3 u BAKER, J. M., and BLEANEY, B.
Bo1 u BOGLE, G. S., and COOKE, A. H.
Bo2 u BOGLE, G. S., and OWEN, J.
Bo3 u BOWERS, K. D.
Ge u GERRITSEN, H. J., BOLGER, B., and OKKES, R. F.
Gr u GRIFFITHS, J. H. E., and WARD, I. M.
Ow1 u OWEN, J.
Ow2 u OWEN, J., and WARD, I. M.
Pa u PALMA, M. U.
Tr u TRENAM, R. S.

THE IONIZATION AND DISSOCIATION OF COMPLEX MOLECULES BY ELECTRON IMPACT

By J. D. CRAGGS† AND C. A. McDOWELL‡

† Department of Electrical Engineering

‡ Department of Inorganic and Physical Chemistry

University of Liverpool

CONTENTS

	PAGE
§ 1. Introduction	374
§ 2. Experimental techniques	376
§ 3. Applications to the study of molecular structure	385
3.1. General introduction	385
3.2. Molecular ionization potentials	389
3.3. Ionization potentials of free radicals	392
3.4. Bond dissociation energies	392
3.4.1. Simple molecules	396
3.4.2. More complex molecules	398
3.5. Collision processes involving negative ions	401
§ 4. Theoretical studies	418
Acknowledgments	420
References	420

Abstract. The report deals with certain experimental and theoretical aspects of collision processes in polyatomic molecular gases. Experimental techniques, dating from about 1932, are reviewed, with more attention given to recent methods. Measurements of critical potentials, cross sections, and initial kinetic energy of product fragments are described.

A survey of the theoretical implications of these data is then given, with reference to the significance of electron-impact ionization potential measurements, for molecules and free radicals, bond dissociation energies for a wide range of molecular types, and certain collision processes involving negative ions.

§ 1. Introduction

An elementary and qualitative picture of ionization processes in molecules is obtained from a consideration of transitions between potential energy curves for the ground state of a molecule XY and for the molecular ion XY^+. The potential energy for the stable state is at a minimum for the equilibrium separation of X and Y and the limiting energy, for infinite inter-atomic separation, corresponds to dissociation of the molecule into fragments X and Y. At the equilibrium position there will be vibrational oscillations with the amplitude shown in figure 1 (taken from Marriott 1954) and the Franck–Condon principle, which postulates that an electronic transition occurs in a time negligible compared with that required for a molecular vibration, thus requires that the inter-nuclear separation will be virtually constant during an electronic transition. In the ground state the eigenfunction $|\psi(R)^2|$ shown in figure 1 (A) indicates that the most probable transition occurs from the centre of the Franck–Condon region, for electron energy E, and in this figure the equilibrium inter-nuclear separations for the ground state of the

initial molecules and of the ion are close enough for electron induced transitions to occur. In other cases this may not be so, and the electron impact and spectroscopic ionization potentials will then differ since the latter values refer to transitions between the lowest vibrational levels of the various electronic states of the molecule

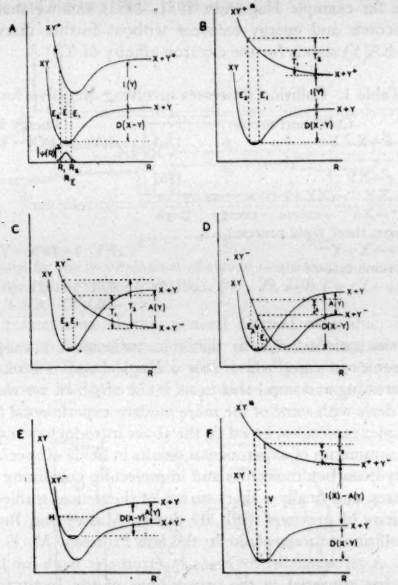


Figure 1. Potential energy curves illustrating ion formation.

and ion whilst the electron impact ionization potentials are essentially vertical on the Franck–Condon diagrams (see also § 3.2). Thus the molecular ion XY^+ is produced as a result of transitions between the two potential energy curves of figure 1 (A), and the reaction may be written simply (e=an electron)

$$XY+e \rightarrow XY^+ + 2e. \qquad \ldots\ldots(1)$$

The process

$$XY+e \rightarrow X+Y^+ + 2e \qquad \ldots\ldots(2)$$

is illustrated by figure 1 (B) where the transition upwards from the Franck–Condon region of the ground state of XY results in the formation of XY^{+*} (an excited state of XY) which dissociates. If $A(Y^+)$ is the so-called appearance potential of Y^+ (i.e. the ionization potential of XY giving Y^+ in this manner) then

$$A(Y^+)=E_1=D(X-Y)+I(Y)+K+E \qquad \ldots\ldots(3)$$

where $D(X-Y)$ represents the dissociation energy of XY, $I(Y)$ is the ionization

potential of Y, K is the total kinetic energy with which the fragments X and Y separate after the collision and E represents excitation energy of the fragments. These processes are discussed in more detail in § 3.

The remaining diagrams in figure 1 represent processes involving negative ion formation. These have been discussed in principle elsewhere by Massey (1950) and others (see for example Hagstrum 1951, 1953) and we shall put down the appropriate reactions and energy balances without further discussion. (In the following table EA(Y) stands for the electron affinity of Y.)

Table 1. Collision Processes involving Negative Ions

Figure	Collision Process	Energy Balance
1(C)	$XY+e \rightarrow X+Y^-$ .....(3*a*)	$A(Y^-)=D(X-Y)-EA(Y)+K+E$ .....(4)
1(D)	$XY+e \rightarrow XY^{-*}$ .....(3*b*)	
E_1 to V	Then $XY^{-*} \rightarrow XY+e$ (low pressure) or $XY^{-*} \rightarrow XY^-+$excess energy (high pressure, three body process)	
V to E_2	$XY+e \rightarrow X+Y^-$	$A(Y^-)=D(X-Y)-EA(Y)$
1(E)	an extreme case of (d)	
1(F)	$XY+e \rightarrow X^++Y^-+e$ .....(5)	$A(X^+)=A(Y^-)=D(X-Y)+I(X)-EA(Y)+K+E$ .....(6)

The above discussions refer to diatomic molecules or, approximately, to polyatomic molecules in cases where only a single bond is broken. The latter covers many interesting and important cases, some of which are mentioned below.

This Report deals with some of the more modern experimental techniques used in studies of complex molecules, based on the above introductory models. Further sections include summaries of experimental results in fields of special interest, such as processes in hydrocarbon molecules and in molecules containing electronegative atoms or complexes, and finally a short survey of theoretical studies is given.

Relevant reviews of previous work are due to Massey and Burhop (1952) in their work on collision processes and to the late Professor M. G. Evans (1952), amongst others. A reference list for mass spectrometer work for January 1938 to November 1950 has appeared in the proceedings of the Institute of Petroleum Conference on Mass Spectrometry (1952) and was prepared by the Metropolitan-Vickers Research Department. Many recent references are given by Barnard (1953), in the National Bureau of Standards publication on *Mass Spectroscopy in Physics Research* (1953) and in the proceedings of the second Institute of Petroleum Conference held in November 1953 (Institute of Petroleum 1954).

§ 2. Experimental Techniques

Early measurements of ionization potentials have been reviewed briefly by Arnot (1933) and more recently and fully by Massey and Burhop (1952) in their critical and comprehensive work. Figure 2 (Tate and Smith 1932) shows an apparatus using collimation of the electron beam by a magnetic field (Jones 1927) suitable for measurements of ionization cross sections and critical ionization potentials. The electrons are accelerated from the filament F through a collimator system into the ionization space between electrodes P_1 and P_2. Ionization produced here is measured as a current to P_1 and P_2 and the onset of this current gives the first ionization potential. Cross sections for ionization are determined

from the magnitude of the ionization current, at any given electron energy, the dimensions of the ionization space, and the gas pressure therein. The magnetic field prevents scattered electrons from reaching the collector and also reduces effects due to secondary emission. Corrections have sometimes to be made for the increase in electron path length in the ionization space due to spiralling along the lines of force (see Massey and Burhop 1952, p. 34 *et seq.*).

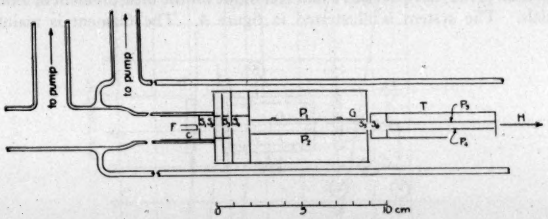


Figure 2. Apparatus for the determination of critical (ionization) potentials and ionization cross sections. (Reproduced from *Phys. Rev.*, **39**, 270).

Subsequent refinements of this general method were due, for example, to Lawrence (1926) and Nottingham (1939). The latter's apparatus for velocity selection of the initial electrons, which will normally possess a fairly wide energy distribution, is shown in figure 3. This enabled more accurate measurements of

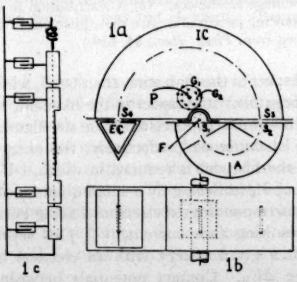


Figure 3. Nottingham's velocity selector. (*a*) Top view showing cathode F; analyser A; ion chamber IC; outer and inner grids G_1 and G_2; ion collector P; and electron cage EC. (*b*) Front view showing location of cathode assembly and slits. (*c*) Side view of cathode assembly only. (Reproduced from *Phys. Rev.*, **55**, 203.)

ionization potentials and cross sections near the critical potentials. These will not be discussed further in this brief review, and we shall proceed to describe the later techniques although so far (in the literature) the methods have only had restricted application to complex molecules.

Thus, detailed studies of the effects of initial electron velocity spread have been made by Vought (1947) and, in more direct experiments by Waldron and Wood (1952), Stevenson (1951) and many others. From the work of Waldron and Wood

on the effect of energy spread in the electron beam, it is clear that for the most precise measurements velocity selection is desirable, although for much technical work such refinements may be unnecessary. Exhaustive reviews of methods of measuring critical potentials in the latter cases are due to Barnard (1953). Hagstrum (1953) has also discussed the problems generally, with reference to the effect of the shapes of the Franck–Condon curves (see § 1 above and Vought (1947)).

Fox *et al.* (1951) have devised a new technique for the measurement of ionization potentials. The system is illustrated in figure 4. The filament is maintained

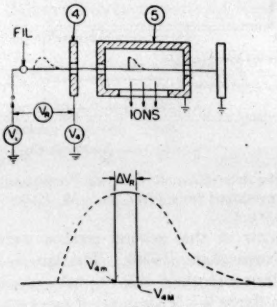


Figure 4. Ion source due to Fox *et al.* (*a*) A sketch of the ion source showing the electrode structure and the voltage references. (*b*) A distribution in electron energy which is modified by the retarding potential to give the equivalent of a monoenergetic electron beam. (Reproduced from *Phys. Rev.*, **84**, 859.)

V_i volts negative with respect to the ionization chamber 5, which is ideally field-free. If now a potential V_4, negative with respect to the filament, is applied to the intermediate electrode 4, then low energy electrons from the filament are prevented from entering the ionization chamber and, effectively, the energy distribution of the electrons emitted from the filament is reduced in width. If the difference in ion currents for two values of V_4 (differing by a small quantity ΔV_R) is determined, it represents the current corresponding to electrons having energies lying within the range ΔV_R and the resulting (ion current, V_4) plot represents the ionization probability measured at a given energy with an electron beam mono-energetic within an energy range ΔV_R. Contact potentials between electrodes 4 and 5 (figure 4) were reduced or eliminated by gold plating; contact potentials between the filament and the accelerating electrodes are irrelevant since the energy scale is set by the potential difference between electrodes 4 and 5.

More work with various forms of velocity selector will doubtless be carried out, and will enable comparisons with this method to be made.

In work with mass spectrometers (which are essential auxiliaries to ionization chambers for measurements on cross sections and appearance (ionization) potentials with complex molecules) further problems arise due to field penetration from the accelerating region into the ionization space. To eliminate field penetration in the ionization chamber Fox *et al.* (1951) introduced a new method, in which the electron current and the ion extraction voltage are pulsed alternately so that, when the

electrons reach the ionization chamber, the extractor field is zero. Thus electrode 5 is a field-free region during the time of formation of ions in it.

Some results are shown in figure 5. Curves (1) and (2) show results taken with

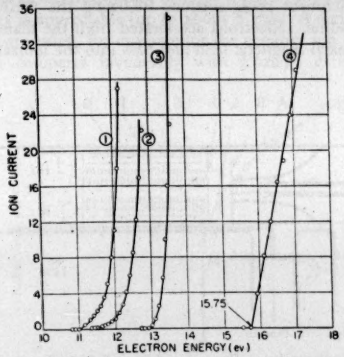


Figure 5. Experimental data obtained with the apparatus of Figure 4. Ionization probability curves for argon 40^+. Run (1) taken by conventional method. Run (2) taken by conventional method with 0.1 sensitivity at the amplifier. Run (3) the ion difference curve using a retarding potential on the electrons, but with an ion-draw-out voltage of 3 volts. Run (4) the ion difference curve obtained using both a retardation of the electrons and pulsed fields on ions and electrons. (Reproduced from *Phys. Rev.*, **84,** 859.)

the conventional method, which involves a finite initial electron energy spread, contact potential differences and a finite extractor field in the ionization chamber. Curve (3) is a curve taken with finite extractor field but with the effects of contact potential differences and initial energy spread removed. Finally, curve (4) shows the results obtained when, with respect to the data of curve (3), the extractor field is removed by the pulsing technique. This new technique appears to hold great promise for mass spectrometer studies, where accurate ionization potentials and cross section measurement are to be made in combination with mass discrimination. Detailed results for Kr and Xe have been obtained by the above authors (Fox *et al.* 1953) with this method.

The molecular processes expressed in table 1, reaction (3) etc., show that information on the kinetic energy term K should often be obtained (exceptions being the study of primary molecular ions such as N_2^+ or, in more complicated examples, $SiCl_4^+$ or $GeCl_4^+$, see § 3, where the term is negligible). The kinetic energy of the ions will appear as a line width in the mass spectrum and it would seem that this would lead to simple kinetic energy measurements. Massey (1950) has reviewed the early work in this field, and later work is due to McDowell and Warren (1951). The instrumental difficulties are appreciable and more definite results have recently been obtained by Hagstrum (1951, 1953) using a modified mass spectrometer, and by Marriott, Thorburn and Craggs (1954) using the method developed by Lozier (1934, see Marriott and Craggs (1954a)).

The apparatus built by Marriott (1954) and used successfully by him and Thorburn in these laboratories, is shown in figure 6; it differs only in trivial details from that used some 20 years ago by Lozier, and was built primarily to give information on initial kinetic energy in negative ion reactions in complex gases and also for work on electron attachment cross sections following the preliminary work of Buchdahl (1941) on [illegible]dine. Electrons accelerated from the filament A are collimated by an external axial magnetic field and pass into the ionization space along

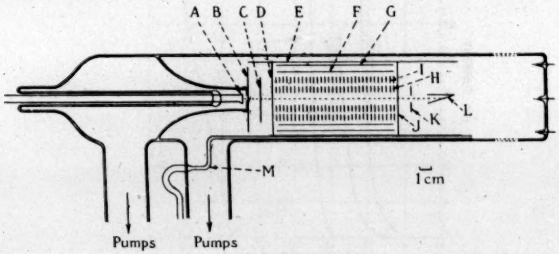


Figure 6. Lozier ionization tube. (Reproduced from E.R.A. Report L/T 308.)

the axis of the electrode H. Ions formed here pass through H and are accelerated to I, another series of coaxial disc electrodes, by a positive or negative voltage depending on the sign of the required ions, and pass through to the collecting electrode F to which a suitable variable voltage is applied to retard the ions passing through I and thus to measure their kinetic energies. The electrodes H are joined to electrodes D, E and J, so that the collisions take place in a field-free region. The circuits are arranged so that the accelerating voltage applied to I is compensated by a reverse voltage on the collecting cylinder.

The kinetic energy analysis is illustrated by figure 7 which shows calibration data, obtained by Marriott, with oxygen. Figure 7 (*a*) shows the (ion current, retarding potential) curve for positive oxygen ions at 20 ev electron energy where the

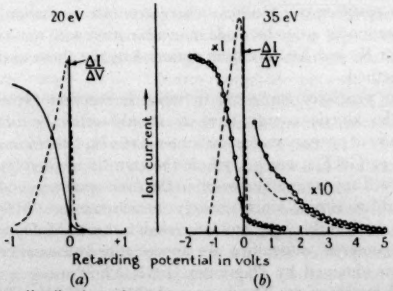


Figure 7. Energy distribution of positive ion from oxygen. (Reproduced from E.R.A. Report L/T 308.)

molecular ion O_2^+ predominates. The ion current cuts off with a very small positive retarding potential, showing that O_2^+ is formed with negligible kinetic energy: indeed the current obtained above zero retarding potential is probably instrumental. At an electron energy of 35 ev (figure 7 (*b*)) O^+ ions are formed and the kinetic energy is appreciable (up to about 5 ev). A series of curves similar to those of figure 7 are taken, for example for the negative ion O^-, and together give figure 8, which compares favourably with Lozier's earlier work. Marriott's

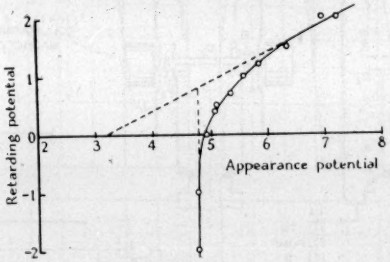


Figure 8. Maximum kinetic energy of negative ions from oxygen-electron energy. (Reproduced from E.R.A. Report L/T 308.)

ionization efficiency curve for O^- is shown in figure 9, and is used, with similar

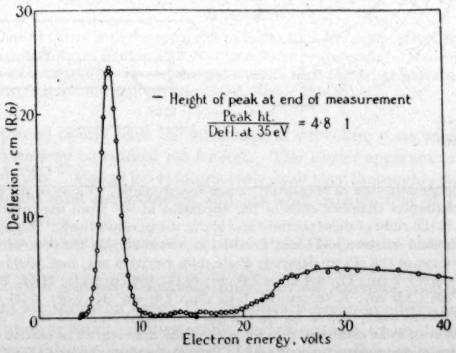


Figure 9. Negative ions from oxygen (O^- from O_2). (Reproduced from E.R.A. Report L/T 308.)

curves from other relatively simple molecules, for calibration purposes in the work on more complex molecules. Some of the later results are shown in § 3 below. The Lozier apparatus does not give mass discrimination and preliminary work with a mass spectrometer to identify the ions is necessary. Even then, if it is shown that the positive or negative ion spectra are complex, i.e. if there is more than one

important positive or negative ion, the Lozier method cannot easily be used. Thus, usually, only the positive or negative ion with the lowest appearance potential can be studied (see § 3 below).

Hagstrum's (1951) apparatus for the measurement of the kinetic energy of products of ionization processes is shown in figure 10 and is essentially a mass

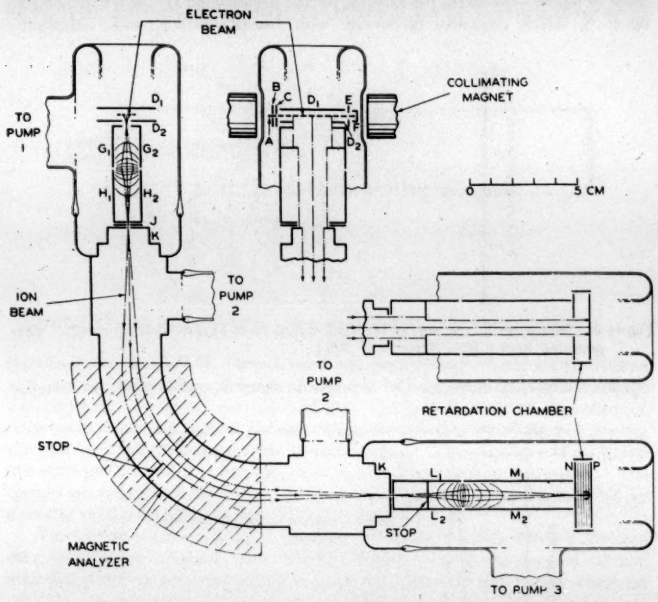


Figure 10. Schematic view of Hagstrum's mass spectrometer. Views of the electron beam and retardation chamber ends of the apparatus at 90° from the principal view are shown to the right of these portions and above them, respectively. Equipotential lines are indicated in the G—H and L—M ion lenses and in the retardation chamber. Dimensions of the slit in electrode B are 0·08 cm × 1·0 cm; in C, 0·03 cm × 1·0 cm; E, 0·1 cm × 1·2 cm; D_2, 0·4 cm × 2·0 cm; G_1G_2, 0·4 cm wide; H_1H_2, 0·05 cm wide; J, 0·1 cm × 2·0 cm; K (entrance), 0·02 cm × 1·0 cm; K (exit), 0·03 cm × 1·0 cm; L_1L_2 0·05 cm wide; stop in L_1 L_2, 0·1 cm × 0·8 cm; N, 0·025 cm × 3·0 cm. Other dimensions may be estimated fom the scale of the drawing. The shaded area indicates the extent of the analyser magnetic field. Typical electrode potentials for $K(e^-) = 100$ ev follow: A, −100 volts; B, −80v; C, D_1, D_2, and E, zero; F, + 45 v; G_1-G_2, −15 v; H_1-H_2, J, and K, −100 v; L_1-L_2, −150 v; M_1-M_2 and N, −30 v; P, −2 → +10 v. (Reproduced from *Rev. Mod. Phys.*, **23**, 185.)

spectrometer with a retarding electrode system. This is arranged so that after acceleration the ions (positive or negative) are returned to the potential, nominally zero, of the ionization space between D_1 and D_2. At this zero potential the energy spread of the ions, due to the initial kinetic energy, may be found by gradually

increasing the potential of P and noting the variation of current at P with the potential. At any setting of the potentials, ions having kinetic energies greater than the retarding field between N and P will be collected at P. Results are shown in figures 11, 12 and 13. This, and earlier similar methods (see Massey and Burhop

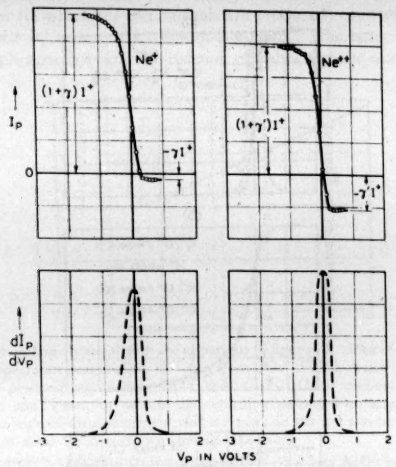


Figure 11. Data obtained with the apparatus of figure 10. Retarding potential curves (top) and their derivatives (bottom) for Ne^+ and Ne^{2+} representing ions formed essentially at rest. Current of positively charged carriers to P plotted as positive values above zero axis. (Reproduced from *Rev. Mod. Phys.*, **23**, 185).

1952 for a review) clearly have the advantage of providing mass analysis, so that complex spectra may be studied ion by ion. The Lozier apparatus above cannot be used in this way. Craig (1954) has recently dealt very thoroughly with retarding slit techniques in mass spectrometry, and his work extends that of Hagstrum, at present under review.

Hagstrum's calibration data are of interest (cf. the results obtained with the Lozier apparatus) and (ion current, retarding potential) (I_p, V_p) curves for Ne^+ and Ne^{2+} (which will possess essentially only thermal energies, negligible in the present work) are shown in figure 11. Secondary emission effects are shown for positive retarding voltages. Other results, including cases where the initial kinetic energy is appreciable, are shown in figure 12 and the (retarding potential, appearance potential) curve for O^+ from CO is given in figure 13. Figure 14 shows (ion current, electron energy) curves for O^- from CO, NO and O_2 taken by Hagstrum. The last curve (O^- from O_2) should be compared with figure 9. We have included these figures to illustrate the influence of instrumental effects on the apparent relative probabilities of negative ion formation over an appreciable electron energy range (figure 9, cf. figure 14, O_2 case). The Lozier apparatus, which does not suffer from field penetration effects, should, we believe, be more reliable in this

respect, although appearance potentials measured with the two kinds of apparatus should agree well after appropriate calibrations have been made.

Many forms of mass spectrometer source have been described, for example, in the reference lists mentioned in § 1 and they will not be treated here. The

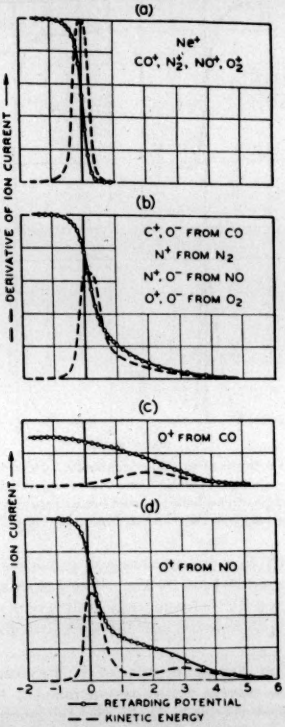


Figure 12. Data obtained with the apparatus of figure 10. Retarding potential and derivative curves for ions formed at rest [part (a) of the figure]; and of three distinguishable types [parts (b), (c), and (d)] for ions from dissociative ionization by electron impact in the gases CO, N_2, NO, and O_2. The curves were taken at $K(e^-)=100$ ev and have been normalized to convenient ordinate scales which bear no meaningful relation to one another. In parts (a) and (b) the experimental curves are for the first ion listed. (Reproduced from *Rev. Mod. Phys.*, **23,** 185.)

Hagstrum ion source, shown in figure 10, is in many respects fairly typical except that most simple sources use only one slit or diaphragm between the filament and the ionization chamber case. In view, however, of the increasing topical interest

in free radicals, it seems desirable to mention the special sources needed for such studies in slightly more detail.

Eltenton's pioneer work (1942, 1947) was intended to explore in a preliminary semi-quantitative manner, a number of chemical reactions by directly detecting the free radicals formed in them. Essentially and ideally the method involves a reaction chamber (in which the free radicals are produced) connected to a suitable mass spectrometer by means of a collision-free path. The latter was realized by the use of a diaphragm, between the reaction chamber and the mass spectrometer,

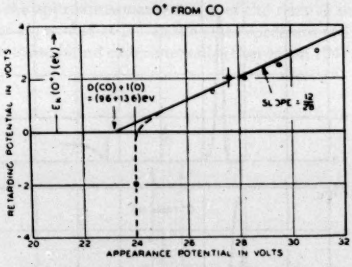


Figure 13. Further data obtained with the apparatus of figure 10. Plot of retarding potential against appearance potential for O^+ ions from CO. Filled circles at $+2$ and -2 volts retarding potential show data from this work. Open circles represent data of Hagstrum and Tate, for which the retarding potential scale is replaced by the equivalent scale of displacement across the peak on the ion accelerating voltage scale. The line of slope 12/28 is drawn through the point

$$[A(O^+)=D(CO)+I(O)=(9{\cdot}6+13{\cdot}6)\ \text{v}=23{\cdot}2\ \text{v};\ K(O^+)=0]$$

in the K, A portion of the graph. (Reproduced from *Rev. Mod. Phys.*, **23**, 185.) *Note.*—E_k in the graph is our K.

and differential pumping. With a pumping speed at the diaphragm of about 7 litres/sec and a reaction pressure of about 100 mm Hg it was possible to maintain a pressure of about 10^{-3} mm in the ionization chamber. Eltenton (1947) describes various reaction chamber units, one of which is shown in figure 15. This apparatus involves a double walled silica cylinder Q, containing a heater. The radicals are produced under the inner wall of Q and pass through D, S_1 and S_2 to the ionization chamber. With apparatus of this kind Eltenton investigated the thermal decomposition of methane, ethane and propylene. The results of later work are discussed in § 3. Lossing, Ingold and Henderson (1953, and elsewhere, see § 3) have developed the method and have applied it to many molecules. One of their reactors is described in which all surfaces with which the reacting molecules collide are composed of quartz. The aperture through which the reaction products pass to the ionization chamber had a diameter of 30 μ and a length of 10–15 μ. Robertson's work (1952, 1954) is also noteworthy.

§ 3. Applications to the Study of Molecular Structure

3.1. *General Introduction*

The ions which are formed by the impact of electrons with molecules arise through processes which obey the Franck–Condon principle as discussed briefly in

§ 1. As the energy of the ionizing electrons is increased, electron transitions occur from the ground state of the neutral molecules to yield excited states of the molecular ions. If some of these excited states are unstable, dissociation processes may occur which produce ions of mass less than that of the parent molecular ion (see § 1). That this view is correct was shown by the studies of the electron induced dissociation of hydrogen by Bleakney, Condon and Smith (1937). The electronic structure of the ground state of molecular hydrogen is given by the molecular

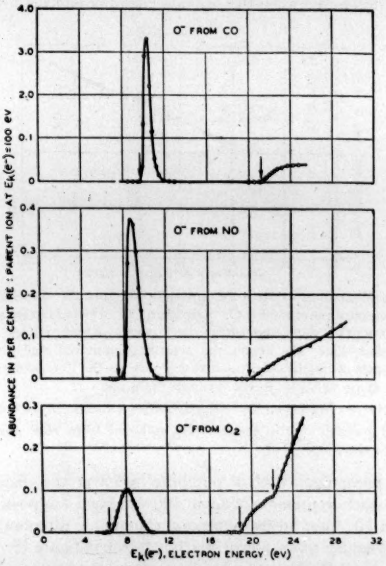


Figure 14. Further data obtained with the apparatus of figure 10. Efficiency of production of O^- ions from CO, NO and O_2 as a function of bombarding electron energy. The data at higher electron energies show the onset of processes in which a positive and negative ion are simultaneously produced [process (5) of table 1]. The so-called resonance peaks at lower electron energies indicate the formation of ions in the capture processes [(3b) of table 1] in these molecules. Note that kinetic energy measurement is essential to the interpretation of these data and the locating of dissociation limits from them. (Reproduced from *Rev. Mod. Phys.*, **23**, 185.) *Note.*—E_k in the graph is our K.

orbital formula $(1s)^2$ and so is a $^1\Sigma_g^+$ state. Spectroscopic data (Beutler and Jünger 1936) yield a value of 15·422 v for the ionization potential of H_2, while electron impact studies lead to a value of $15{\cdot}4 \pm 0{\cdot}1$ v (Bleakney 1932, Smith and Bleakney 1936, Hagstrum and Tate 1941). This gives the energy required to remove an electron to form the H_2^+ ion in the ground $^2\Sigma_g^+$ state. Above this $^2\Sigma_g^+$ state to H_2^+ there lies the $^2\Sigma_u^+$ state which has the electronic structure $(2\,p\sigma)$. The Wigner–Witmer rules show that both the $^2\Sigma_g^+$ and the $^2\Sigma_u^+$ states

arise from the combination $H(^2S_g)+H^+$. The potential energy curves of H_2 and H_2^+ are well known and it has been found experimentally that there are two modes of formation of H^+ ions as the result of electron impact. These occur at exactly the energies required by the application of the Franck–Condon principle. One group of H^+ ions is formed with near zero kinetic energy by transitions to the dissociation limit of the $^2\Sigma_g^+$ state of H_2^+; the second group is formed with appreciable kinetic energy and these arise by transitions to the repulsive $^2\Sigma_u^+$ state of H_2^+. Here it is interesting to note that the application of the Franck–Condon principle, to permit the approximate calculation of the relative intensities of H^+ in the mass spectrum of H_2 to that of D^+ in the mass spectrum of D_2, gives results in good agreement with those found experimentally (Stevenson 1947, see also Schaeffer and Hastings 1950).

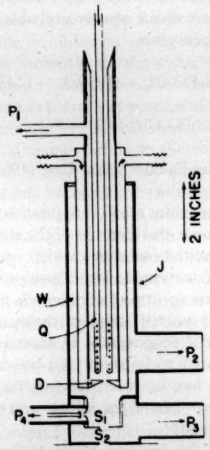


Figure 15. A free-radical source due to Eltenton. Low pressure quartz reactor attached to ionization chamber of mass spectrometer. Ground joint J, water-cooled jacket W, quartz double-walled tube Q, diaphragm D, ion-collimating slits S_1 and S_2, pumping leads P_1, P_2, P_3 and P_4. (Reproduced from *J. Chem. Phys.*, **15**, 455.)

Unfortunately little is known about the potential energy curves or surfaces of many molecular ions, but that the processes leading to the formation of ions as the result of electron impact is the same for complex molecules as for hydrogen, is shown by the observation that metastable ions are frequently produced which dissociate to yield other ions of smaller mass. Thus it has been found (Hipple, Fox and Condon 1946) that in the mass spectrum of *n*-butane there is a broad peak at mass 31·9 which is due to the dissociation of the metastable ion $C_4H_{10}^+$ by the process:

$$C_4H_{10}^+ \rightarrow C_3H_7^+ + CH_3.$$

Many such metastable ions are known and it has been found that they have a half-life of about 10^{-6} second. Examples of metastable molecular negative ions have also recently been observed. Ahearn and Hannay (1953), whose work is discussed in more detail in § 3.5, in examining the negative ions formed when sulphur hexafluoride is subjected to electron impact found that a negative ion was produced with an apparent mass 16·4 mass units lighter than SF_5. This corresponds to the metastable transition $(SF_6^-)^* \rightarrow SF_5^- + F$. Since SF_6^- was also detected it is apparent that both of the following two processes occur when the SF_6 molecule captures an electron:

$$SF_6 + e \rightarrow SF_6^{-*} \begin{cases} \nearrow SF_6^- \\ \searrow SF_5^- + F \end{cases}$$

The peak heights of SF_6^- and SF_5^- were approximately the same, indicating that both modes of dissociation are about equally probable. Other metastable negative ion transitions include the processes

$$C_6H_5PO_2Cl_2^- \rightarrow PO_2Cl^- + C_6H_5Cl$$

$$PO_2Cl^- \rightarrow Cl^- + PO_2$$

which have been observed by Donnolly and Carr (1954) amongst the negative ions formed from $C_2H_5PO_2Cl_2$.

These metastable ion transitions are akin to predissociation phenomena observed in optical spectra but, of course, the lifetimes of the metastable ions are many times greater than those of the excited molecules which undergo predissociation. The experimental observation of metastable ion transitions substantiates the view that the ions observed in the mass spectra of compounds arise as the result of Franck–Condon transitions to yield excited states of molecular ions which dissociate to form smaller ions and neutral fragments. In most cases the lifetime of the dissociating parent ions are extremely short and so cannot be identified. Some progress in theoretical work has, however, recently made it possible to identify the vibronic states of the parent ions from which certain smaller ions are formed Laidler 1954, McDowell 1954b).

The potential at which the ionizing electrons have just sufficient energy to cause a particular ion to appear in the mass spectrum of a compound is known as the appearance potential and the accurate determination of this quantity for a large number of ions is of great importance. Earlier in this article (§ 2) we have discussed the various experimental methods for measuring appearance potentials of both positive and negative ions. Since the ions arise as the result of Franck–Condon transitions, it follows that the appearance potentials of parent molecular ions are the vertical ionization potentials of the molecules, and these give information about the electronic structure of the molecules. In the case of ions formed by dissociation processes, appearance potentials can yield important information about the dissociation energies of chemical bonds, the ionization potentials of free radicals, and the electron affinities of atoms and molecules. As all these quantities are of great importance in the complete understanding of the electronic structure of molecules, we shall now discuss how such information may be obtained from electron impact studies.

3.2. *Molecular Ionization Potentials*

Besides the electron impact method the ionization potentials of molecules can be obtained by studying that part of the far ultra-violet spectrum which is due to Rydberg transitions. Once the Rydberg transitions have been identified it is possible to fit them to a series and so to calculate the ionization potential of the molecule. In favourable cases this method yields highly accurate values for molecular ionization potentials. There are often difficulties in the interpretations of the spectra and then an unambiguous assignment of the Rydberg bands is not possible. Electron impact studies will, however, always yield a value for the ionization potentials of molecules except in cases such as CF_4 and SiF_6 where the parent ions CF_4^+ and SiF_6^+ are so unstable that they cannot be detected in sufficient abundance. It will, of course, be realized that the spectroscopic data are adiabatic ionization potentials, that is, they refer to the production of the ion in its equilibrium configuration whereas electron impact studies yield values for the vertical ionization potentials (see § 1); these latter are generally a little larger (Honig 1948, Morrison and Nicholson 1952). If the equilibrium configuration and dimensions of the molecular ion and the parent molecule are the same, then both the spectroscopic and the electron impact ionization potentials will be equal within experimental error.

The spectroscopic method has been most successful when applied to unsaturated molecules which have ionization potentials between about 8·5 and 11·5 v. In this region the ionization potential usually refers to the energy required to remove an electron from a π molecular orbital. The far ultra-violet spectra of molecules containing carbon–carbon bonds becomes increasingly difficult to interpret in the region where the energy of the electronic transitions is greater than about 11·5 v. Here transitions due to σ-bonding electrons begin to become important and the spectra are frequently diffuse. It is for this reason that there are no spectroscopic values for the ionization potentials of saturated hydrocarbons though accurate values for these are available from electron impact studies (Honig 1948). Because of experimental difficulties of a different nature, there are also no spectroscopic values available for the ionization potentials of free radicals though such data have been obtained by electron impact methods (Hipple and Stevenson 1943, Lossing, Ingold and Henderson 1954a, b, LeGoff 1953).

The ionization potential of a molecule gives the energy required to remove an electron from the highest occupied molecular orbital and so such data help one to understand the electronic structures of molecules. The fact that generally only the first ionization potentials of molecules are available is a great handicap but such data have, however, enabled some important theoretical concepts to be assessed. For example, the red shifts in the 1800 Å bands of ethylene, which accompanies alkyl substitution has been explained in terms of hyperconjugation (Mulliken, Rieke and Brown 1941). This energy change is paralleled by the change in the ionization potential from ethylene (I=10·6 v), to propylene (I=9·84 v). Many suggestions have been put forward to account for this decrease in the ionization potential as a result of alkyl substitution, but as there are many different factors involved such as inductive effects, change of bond hydridization, and hyperconjugation, it is difficult accurately to assess the importance of each of these. Recently Hall (1953) has shown that the observed decrease can also be explained in terms of the interaction between electrons in *equivalent* orbitals. A similar lowering of the ionization potential of the non-bonding $p\pi$ electrons of the oxygen atom in the water molecule as the result of alkyl substitution has also been observed (Cox and

McDowell 1953, unpublished). This is illustrated by the data in table 2.

Table 2. Ionization Potentials of Water, Alcohols and Ethers

Molecule	H_2O	CH_3OH	C_2H_5OH	n–C_3H_7OH	$(CH_3)_2O$	$CH_3OC_2H_5$	$(C_2H_5)_2O$
Ionization potential (ev)	12·58	10·86	10·39	10·22	10·13	9·81	9·62

The ionization potentials of several isotopic molecules have been studied by electron impact methods and the results are given in table 3.

Table 3. Ionization Potentials of Isotopically Substituted Molecules (ev)

(*a*)		(*b*)	(*c*)	(*c*)		(*d*)	
H_2O	12·58±0·03,	CH_4 13·04±0·02,	13·18±0·02,	C_2H_2	11·40±0·02,	NH_3	11·30±0·02
HOD	12·58±0·03,	CH_3D —	13·21±0·02,	C_2D_2	11·39±0·02,	ND_3H	11·47±0·02
D_2O	12·60±0·02,	CH_2D_2 —	13·25±0·02,	C_2H_4	10·60±0·02,	ND_3	11·52±0·03
		CHD_3 —	13·32±0·02,	C_2D_4	10·59±0·02,	(all on arbitrary	
		CD_4 13·21±0·02,	13·36±0·02,			scale)	

References (*a*) Cox and McDowell 1953, unpublished experiments.
(*b*) Honig 1948.
(*c*) Lossing, Tickner and Bryce 1951.
(*d*) Neuert 1952.

The large difference in the ionization potentials of CH_4 and CD_4 and NH_3 and ND_3 is most unexpected. It would have been expected that the only difference would have been that due to the change in the zero point energies of the neutral molecule and the parent molecular ion. Estimates of this difference in the zero point energies for CH_4–CH_4^+ and CD_4–CD_4^+ shows that this does not account for the observed differences in the ionization potentials (Lossing, Tickner and Bryce 1951). It is interesting to note that the ionization potentials of the free radicals CD_3 and CH_3 have been found to be the same within 0·03ev (Lossing, Ingold and Henderson 1954b).

Though the information which can be obtained from a knowledge of only the first ionization potential of a molecule is somewhat limited, such data have, however, proved of use in providing assessments of molecular orbital theories. The view has been taken that they provide a more reliable estimate of the accuracy of molecular orbital calculations than a comparison of the predicted and experimental absorption spectra. Example of this type are the cases of ethylene (Mulliken 1949) and allene (Parr and Taylor 1951), water (Ellison and Shull 1953), and several unsaturated compounds (Pople 1953). Hall (1951, 1953) has proposed a new semi-empirical method in terms of equivalent orbits (Lennard-Jones 1949, Hall and Lennard-Jones 1950). From this theory Hall has calculated the ionization potentials of the saturated hydrocarbons from propane to decane and obtains good agreement with Honig's (1948) experimental values. In this work it was assumed, as had been suggested somewhat earlier by Mulliken (1935 a) that the first ionization potential referred to the removal of an electron from the $[\sigma a_1]$ symmetrical molecular orbital mainly localized over the C–C bonds of the molecule. If it is assumed that the ionization potential arises from the removal of an electron from an anti-symmetrical orbital localized over the C–H bonds, Hall (1954) has shown that just as good agreement

between the calculated and the experimental values is obtained. The ionization potentials of some derivatives of the paraffins have also been considered in an attempt to obtain some decisive evidence indirectly but no unequivocal choice can be made. It appears that the problem will only be solved when all the ionization potentials of some hydrocarbons or their derivatives have been determined and analysed (see below). In this connection it may be noted that recent measurements of the ionization potentials of the alkyl halides (McDowell and Cox 1954) have provided support for the molecular orbital formulae for these compounds given by Mulliken (1935 b).

The great need for information about the inner ionization potentials of molecules has been stressed above and though a little information is available most of this has been obtained by the use of simple ionization tubes without mass analysis and so the ionization process could not be identified with certainty. Recently, however, the development of an essentially mono-energetic source of electrons with an energy spread of 0·1 ev by Fox *et al.* (see below) has made it possible to measure the inner ionization potentials of several molecules.

The only recent work using the simple ionization tube method is that of Sugden and Price (1948), who studied the ionization of water, hydrogen sulphide, formaldehyde, acetaldehyde and acetone. The lack of mass analysis, as has been mentioned before, makes it difficult to distinguish simple ionization processes, and so it is difficult to say whether this type of work is of great value. In the case of the water molecule, however, the observed ionization potentials agree fairly well with those to be expected from recent molecular orbital calculations (Ellison and Shull 1953), but more work is clearly necessary before the proposed assignments of the various processes can be accepted.

We have earlier mentioned that developments in the various molecular orbital theories of the electronic structures of molecules have focused attention on the need for information about the several ionization potentials of molecules and the Fox–Hickam–Kjeldaas–Grove method for obtaining essentially mono-energetic electrons has now been used to study the ionization of several molecules (Frost and McDowell 1954, unpublished). In the case of nitrogen there have been found all the four ionization potentials involving the formation of the N_2^+ ion in its four known excited states, namely, $X^2\Sigma_g^+$, $A^2\Pi_u$, $B^2\Sigma_u^+$, and the $C^2\Sigma_u^+$. The newly determined ionization potentials are in excellent agreement with those to be expected from spectroscopic data. The methyl halides have also been studied and here it has been possible to detect all the excited states of the CH_3X^+ ions to be expected from the molecular orbital formulae of these compounds as proposed by Mulliken (1935b). It is also interesting to note that in the cases of CH_3I and CH_3Br it has been possible to observe the splitting of the ^2E ground levels of the molecular ions, and the values of the doublet separations found agree with those given earlier by Price (1936) from the Rydberg series which fitted the bands in the far ultra-violet spectra of these compounds. Morrison (1953) has described a technique which he has used to try to observe excited states of molecular ions. This method is at present restricted to using electrons with an energy spread of about 0·3 ev and this, of course, places a serious restriction on the accuracy with which the ionization efficiency curves of molecules can be determined. A recent paper (Morrison 1954) describes the results of the application of this method to hydrogen, oxygen, nitric oxide, nitrogen, carbon dioxide, benzene and acetylene. Though the existence of several excited states of molecules is shown, the agreement between the experi-

mentally observed excitation energies and those to be expected from a knowledge of the spectra of the ions leaves much to be desired. In the cases of N_2^+ and CO_2^+ the agreement between theory and experiment is not good, as the experimental peaks are always found to be 0·5 to 1·0 ev above the known levels. Considerable improvements will, therefore, have to be made if this method is to be used in determining the energies of excited states of molecular ions.

3.3. *Ionization Potentials of Free Radicals*

As mentioned earlier, there are no spectroscopic ionization potentials available for free radicals, though recently such data for several radicals have been obtained by electron impact methods. Just as for molecules, these data are of interest in the assessment of theories of the electronic structures of the radicals. They are of great importance in enabling bond dissociation energies to be computed by the direct method (see § 3.4) and are also necessary if the energetics of certain organic chemical reactions in solution are to be understood (Evans 1946).

Free radicals were first detected mass spectrometrically by Eltenton (1942, 1947) as shown briefly in § 2, but the ionization potential of a free radical (methyl) was first measured by this method by Hipple and Stevenson (1943). These authors generated the radicals by the pyrolysis of lead tetramethyl in a small furnace inside the mass spectrometer and allowed the radicals to effuse through a small quartz tube into the ionization chamber. Besides the methyl radical, Hipple and his colleagues (Hipple and Stevenson 1943, Langer and Hipple 1946) were also able to determine the ionization potentials of the ethyl and methylene radicals. Recently Lossing and his colleagues (Lossing and Tickner 1952, Lossing, Ingold and Henderson 1954a, b, Farmer and Lossing 1954, unpublished) have considerably developed this method for obtaining a molecular beam of free radicals and have measured the ionization potentials of several interesting radicals. The ionization potentials of all the free radicals which have thus far been determined are given in table 4.

There are several interesting features displayed by the data in table 4. The decrease in ionization potentials of the radicals CH_3, C_2H_5, C_3H_7 is similar to that earlier noticed with the alkylethylenes. Thus we have:

Molecule	$CH_2{=}CH_2$	$CH_3CH{=}CH_2$	$(CH_3)_2C{=}CH_2$
I(ev)	10·62	9·84	8·8
Radical	CH_3.	CH_3CH_2.	$(CH_3)_2CH$.
I(ev)	9·95	8·78	7·90

A second interesting group is that composed of the allyl, propargyl and the benzyl radicals. All these have a conjugated structure though the allyl radical is the only one which is completely symmetrical. Because of this one would expect that the delocalization energy of the π electrons in the allyl radical would be slightly greater than is the case for the propargyl radical. Thus on these grounds it would be expected that $I(\text{allyl}) < I(\text{propargyl})$ and this is borne out by the experimental data. Again the simple L.C.A.O. molecular orbital theory leads to the conclusion that the delocalization energy of the benzyl radical is slightly greater than that of allyl and the above results show that $I(\text{benzyl}) < I(\text{allyl})$. In these comparisons it must be remembered that part of the decrease in the ionization of the radicals just considered may be due to an increase in the stability of the corresponding ion.

Finally, it is to be noted that the ionization potential of the CCl_3 radical is very much less than that of methyl, and this can be explained on the basis that there is

Table 4. Ionization Potentials of Free Radicals

Radical	Source	Ionization potential (ev)	Authors
methyl (CH_3)	lead tetramethyl	$10{\cdot}07\pm0{\cdot}1$	Hipple and Stevenson (1943)
,,	mercury dimethyl	$9{\cdot}95\pm0{\cdot}03$	Lossing, Ingold and Henderson (1954)
,,	,, ,,	$10{\cdot}11\pm0{\cdot}01$ $10{\cdot}04\pm0{\cdot}12$	Waldron (1954)
,,	lead tetramethyl	$9{\cdot}85\pm0{\cdot}2$	LeGoff (1953)
deuteromethyl (CD_3)	mercury di-deuteromethyl	$9{\cdot}96\pm0{\cdot}03$	Lossing, Ingold and Henderson (1954)
methylene (CH_2)	diazomethane	$11{\cdot}9\ \pm0{\cdot}2$	Langer and Hipple (1946)
ethyl (C_2H_5)	lead tetra-ethyl	$8{\cdot}67\pm0{\cdot}1$	Hipple and Stevenson (1943)
,,	mercury di-ethyl	$8{\cdot}78\pm0{\cdot}05$	Farmer and Lossing (1954, unpublished)
allyl ($.CH_2{-}C{=}CH_2$)	allyl iodide and *l*-butene	$8{\cdot}16\pm0{\cdot}03$ $8{\cdot}15$	Lossing, Ingold and Henderson (1954)
isopropyl ($(CH_3)_2CH\cdot$)	azo-isopropane	$7{\cdot}90\pm0{\cdot}05$	Farmer and Lossing (1954, unpublished)
propargyl ($.CH_2{-}C{\equiv}(H)$)	propargyl iodide	$8{\cdot}25\pm0{\cdot}08$	Farmer and Lossing (1954, unpublished)
trichloromethyl ($.CCl_3$)	chloropicrin	$8{\cdot}78\pm0{\cdot}05$	Lossing, Farmer and Henderson (1954, unpublished)
benzyl ($C_6H_5CH_2\cdot$)	benzyl iodide	$7{\cdot}73\pm0{\cdot}08$	Lossing, Ingold and Henderson (1954)
	benzylamine	$7{\cdot}81\pm0{\cdot}08$	Farmer, Henderson, McDowell and Lossing (1954)

some interaction between the electrons in the $4p\pi$ non-bonding orbital of the Cl atoms and that in the $2p_z$ orbital of the central carbon atom, and also partially to the greater stability of the CCl_3^+ ion over the CH_3^+ ion.

3.4. *Bond Dissociation Energies*

Modern chemical kinetic studies have shown that most reactions occur by bimolecular processes in which chemical bonds are either formed or ruptured. One of the primary aims of kinetic studies is the determination of the activation energies of these elementary steps and to show how these energies are related to the thermal changes in the processes. To predict such values a knowledge of bond dissociation energies is required. Further, if the theoretical factors influencing the stability of molecules are to be understood fully, then it is again necessary to know the dissociation energies of many bonds, and how these are influenced by their molecular environment. Bond dissociation energies are also of importance in providing data which are often of assistance in elucidating the nature of certain details in optical spectra such as predissociations and related phenomena.

The bond dissociation energy of a particular bond $R_1{-}R_2$ is defined as the energy required to dissociate the molecule R_1R_2 into the radicals R_1 and R_2 in their standard thermochemical states at 0°K. Thus the bond dissociation energy $D(R_1{-}R_2)$ is the heat of the reaction

$$R_1R_2 = R_1 + R_2 - D(R_1{-}R_2)$$

and is given by the equation

$$D(R_1{-}R_2) = \Delta H_f^0(R_1R_2) - \Delta H_f^0(R_1) - \Delta H_f^0(R_2)$$

where $\Delta H_f^0(R_1R_2)$, $\Delta H_f^0(R_1)$, and $\Delta H_f^0(R_2)$ are the heats of formation of the particular species in their standard states at 0°K. It has, however, been customary for physical chemists to neglect the difference between the bond dissociation energies at 0°K and those at 25°C and so in the literature most of the data refer to calculations based on heats of formation in the standard states of the species at 25°C.

There are many methods which have been used to estimate the dissociation energies of certain chemical bonds (Gaydon 1952, Szwarc 1950). These methods depend on a knowledge of such phenomena as (*a*) the chemical kinetics of certain processes in which a particular bond is ruptured, (*b*) photochemistry and (*c*) optical spectra involving phenomena such as predissociations. All of these methods are of importance but the one now to be described which is based on data obtainable from electron impact studies is in principle more generally applicable.

The electron impact method of estimating bond dissociation energies is based on the fact that the appearance potential of an ion formed in a dissociation process is determined by quantities which can be measured or estimated. Thus, if under the influence of electron impact the molecule XY dissociates to form the Y^+ ion by the process (2) of § 1 then the appearance potential of the Y^+ ion $A(Y^+)$ is given by the equation (3), § 1.

If the last three quantities in equation (3) are known, then the dissociation energy of the X–Y bond can be computed from the appearance potential of the Y^+ ion.

The correct assignment of the dissociation process is obviously of primary importance. Thus if the X^+ ions were produced by the process (7)

$$XY+e=X^++Y^- \qquad \ldots\ldots(7)$$

instead of that represented in equation (1) there would have to be included in equation (3) an extra term EA(Y) to represent the electron affinity of Y. Fortunately by studying the formation of both positive and negative ions as a function of the energy of ionizing electrons, it is possible to decide between the two types of dissociation processes. In the case of positive ions formed from molecules containing highly electronegative groups such as CN, OH, Cl, Br, etc., it is particularly important that the negative ion spectra be studied to ensure the accurate identification of the process by which the positive ion is formed. Examples of cases where it has been possible by the above means to identify correctly the dissociation processes are:

$$N_2+e=N^++N$$
$$CO+e=C^++O$$
$$CO+e=C+O^+$$
$$CH_3I+e=CH_3^++I$$
$$CH_3CN+e=CH_3^++CN$$
$$CCl_4+e=CCl_3^++Cl^-.$$

Once the dissociation process has been identified, it is next necessary to know whether the ions are produced with excess kinetic energy and also whether they and the neutral molecular fragments are formed in their ground or in excited states. Various workers have assumed that the ions and molecular fragments produced in certain dissociation processes are formed in excited states. In none of these cases has any experimental evidence been given to show that this is so; rather it has been generally the case that the products have been assumed to be in an excited state so

that the sum of the energies of the various terms on the right-hand side of, for example, equation (3) should equal the appearance potential. Thus in the electron induced dissociations of CO, N_2 and O_2 (Hagstrum and Tate, 1941, Hagstrum 1947 a, b), it has been assumed that in the following processes the neutral atom is produced in an excited state:

$$CO(X^1\Sigma^+) \rightarrow C^+(^3P^0) + O(^1D)$$
$$N_2(X^1\Sigma_g^+) \rightarrow N^+(^3P) + N(^2D^0)$$
$$O_2(X^3\Sigma_g^-) \rightarrow O^+(^4S^0) + O(^1D).$$

Similarly it has been assumed (Stevenson 1951) that when CN^+ ions are formed from cyanogen, the cyanide radical is produced in its $^2\Pi_i$ excited states, i.e.

$$(CN)_2 + e \rightarrow CN^+ + CN(^2\Pi_i).$$

In this latter case it has been shown (McDowell and Warren 1952) that this assumption is without foundation and the excess energy is due to the CN^+ ion being produced with kinetic energy. Though there may undoubtedly be cases in which the neutral fragment is formed in an excited state and, of course, such processes are to be expected on theoretical grounds, there are at present no reasons for thinking that this happens very frequently. In most cases, unless there are compelling reasons for the contrary opinion, it is best to assume that the neutral fragments are formed in their ground states.

There remains to be considered the excess kinetic energy with which the ions may be formed. Perhaps the most satisfactory way of measuring the kinetic energy of an ion produced in a dissociation is by means of the Lozier apparatus which has been described above. Unfortunately, however, this type of apparatus does not have provision for mass analysis and so occasionally even when mass spectral data are available it may be difficult to be certain that one is studying a particularly desired process. There are, however, several mass spectrometric methods for detecting the presence of ions with excess kinetic energy (Bleakney 1930, Fox and Hipple 1948, Berry 1950, McDowell and Warren 1951), and also some for estimating the amount of this excess energy (Hagstrum and Tate 1941, Hagstrum 1947 a, b). The method used by McDowell and Warren (1951) is at present purely qualitative and depends on the fact that there is a certain amount of aberration for beams of ions which are produced with excess kinetic energy, thus such ion beams have a greater half-width than those found for ions which are produced without excess kinetic energy. Hagstrum and Tate (1941) made an extensive study of the shapes of ion beams and were able to distinguish between ions which were formed with excess kinetic energy and those which were not. They were also able to determine the distribution in the initial kinetic energies of the ions. By this method they were able, for example, to show that the O^+ ion from CO is produced with excess kinetic energy and by measuring the appearance potential of this ion when it was produced with 1·5, 2·0, 2·5 and 3·0 ev excess energy, they were able to deduce the appearance potential which referred to the production of the ion with zero kinetic energy. This method has recently been used by Waldron (1954) to show that in the electron induced dissociation of the methyl radical the CH_2^+ and CH^+ ions are produced without excess kinetic energy.

The more recent method used by Hagstrum (1951) is a development of earlier work by Hipple, Fox and Condon (1946), Fox and Hipple (1948) and Berry (1950),

in which the amount of the kinetic energy is measured by using retarding potentials on plates at the exit slit of the mass analyser (see § 2 and figure 10). He has applied this method to estimate the amount of kinetic energy possessed by the ions formed in the dissociation of N_2, H_2, NO, CO and O_2. Having thus mentioned the methods available for detecting and estimating the excess energy which may arise in an electron induced dissociation process, we shall now consider some of the results of using equation (3) to estimate bond dissociation energies from appearance potential measurements.

3.4.1. *Simple molecules.*

These will be discussed briefly as a necessary introduction to the work on complex molecules.

(i) *Hydrogen.* Very complete studies of the ionization and dissociation of hydrogen have been made and it has been shown by Bleakney (1932), Hagstrum and Tate (1941) and Hagstrum (1951) that H^+ ions are formed with near zero kinetic energy at $18{\cdot}0 \pm 0{\cdot}2$ ev. These have also been shown to arise by the process

$$H_2 + e \rightarrow H^+ + H + 2e$$

and so

$$A(H^+) = D(H_2) + I(H).$$

The ionization potential of the hydrogen atom is 13·57 v; hence a value of $4{\cdot}43 \pm 0{\cdot}2$ ev is obtained for the dissociation energy of the hydrogen molecule which is in excellent agreement with that deduced from spectroscopic measurements (Herzberg 1950).

(ii) *Nitrogen.* The electron induced dissociation of nitrogen has been studied by Lozier (1933, 1934), Hagstrum and Tate (1941) and Hagstrum (1951). In this case it has been clearly established that negative ions are not formed, and the N^+ ions arise by a simple process in which N atoms are produced simultaneously. Here it has also been established that N^+ ions are produced with near zero kinetic energy at a potential of $24{\cdot}3 \pm 0{\cdot}2$ v. Hagstrum has given reasons for assuming that in the dissociation process the N atoms are produced in their ^2D_u excited state, and so deduces a value of 7·4 ev for the dissociation energy of the nitrogen molecule. If, however, it is assumed instead that the N atoms are produced in their ^4S_u state, then there results the value of 9·8 ev for $D(N_2)$. On the basis of the above data it is not possible to decide between these two values, but by considering the possible processes in the electron induced dissociation of nitric oxide, Hagstrum concludes that the value of 7·4 ev is the true dissociation energy of nitrogen molecules. Many different values for the dissociation energy of the nitrogen molecule have been suggested by the different interpretations of the optical spectrum of this substance. Of these it now appears that only the two values of 7·37 and 9·75 ev need be considered. The former value has been advocated by Herzberg (1950), Long and Norrish (1946) and Long (1949) and, as has been mentioned, it is strongly supported by Hagstrum's (1951) electron impact studies. The larger value has been vigorously defended by Gaydon and Penney (1945) and Gaydon (1952); and more recently Douglas (1952) has shown that new spectroscopic data on the N_2^+ molecule on the whole favour the high value, though the low one cannot with certainty be eliminated. Other experiments of a different nature by Kistiakowsky, Knight and Malin (1951) lend support to the high value. Very recent mass spectrometric studies by Father Clarke (1954) have shown conclusively that the dissociation process is

$N_2+e \rightarrow N^+(^3P)+N(^4S)$ and so it follows $D(N_2)=9{\cdot}75$ ev.

(iii) *Nitric oxide.* The work of Hagstrum and Tate (1941) and Hagstrum (1951) has established that from this molecule there are produced as the result of electron induced dissociation processes N^+ and O^- ions with near zero kinetic energy and O^+ ions, both with near zero kinetic energy and with a mean kinetic energy of about 3 ev. Using the methods discussed above Hagstrum has been able to determine the appearance potentials for the O^- ion when it is produced without kinetic energy. The appearance potential of the O^+ ions produced without kinetic energy could not be measured and Hagstrum inferred its value. He suggests feasible dissociation processes for the production of the various ions and on the basis of these deduces, as mentioned above, a value of $D(N_2)=7{\cdot}34$ ev. He is also led to a value of 2·2 ev for the electron affinity of the oxygen atom.

(iv) *Carbon monoxide.* Studies of the dissociation of this molecule by Lozier (1934), Hagstrum and Tate (1941) and Hagstrum (1951) have shown that here the C^+ and O^- ions formed with near zero kinetic energy but that the O^+ ion is produced with excess kinetic energy. Possible dissociation processes for these ions have been put forward and those advocated by Hagstrum lead to a value of 9·6 ev for the dissociation energy of the CO molecule. As in the case of N_2, here again it is not possible to give an unequivocal interpretation of the spectrum of this compound which will lead to a unique value for $D(CO)$. Values ranging from 6·92 ev to 11·108 ev have been suggested by various workers on the basis of different interpretations of the spectroscopic data. This controversy has been reviewed in detail by Herzberg (1937, 1950), Gaydon (1952), Hagstrum (1947) and Long (1949), and little new can be added to these accounts.

(v) *Oxygen.* Lozier (1934), Hagstrum and Tate (1941), and Hagstrum (1951) have made extensive studies of the dissociation of oxygen. It has been shown that O^+ and O^- ions with near zero kinetic energy are produced. O^- ions are produced with kinetic energy in a resonance capture process at low energies. The various possible dissociation processes have been fully discussed by Hagstrum and the data have been interpreted to show that the electron affinity of the oxygen atom must be approximately 2·2 ev. The dissociation energy of the oxygen molecule is found to be 5·1 ev which is in very good agreement with that obtained from spectroscopic data. Later work of Thorburn (1954) provides a better interpretation of the oxygen dissociation processes.

(vi) *Hydrogen cyanide, cyanogen, and cyanogen chloride.* Recently Stevenson (1950) has measured the appearance potentials of the CN^+ ion from HCN and $(CN)_2$. On the assumption that this ion is produced without excess kinetic energy, it is possible to combine appropriate thermochemical data with the difference between the two values of the appearance potentials and so deduce a value for the dissociation energy $D(NC–CN)$ of $6{\cdot}88-2E(CN)$ ev, where $E(CN)$ is the excitation energy of the CN radical. The appearance potential of the Cl^+ ion from ClCN was also measured and a similar treatment of the data lead to the result $D(NC–CN)=6{\cdot}92-2E(CN)$. Stevenson has given arguments to support the view that the CN radical is produced in its $A^2\Pi_i$ excited state from which it would follow that $D(NC–CN)=4{\cdot}64\pm0{\cdot}2$ ev. The reasons given for this interpretation are by no means compelling, and more recent work has shown that the CN^+ ion from $(CN)_2$ is produced with excess kinetic energy (McDowell and Warren 1952).

The dissociation energy $D(NC–CN)$ cannot be determined by the above method

unless retarding potential methods are used to measure the amount of excess kinetic energy which the CN^+ ion possesses. D(NC–CN) can, however, be determined indirectly from the results of electron impact experiments on methyl cyanide CH_3CN. It has been shown (McDowell and Warren 1952) that the CH_3^+ ion is produced without excess kinetic energy from this compound, and the combination of electron impact results with thermochemical data enables the heat of formation of the CN radical to be calculated. The value obtained in this way is in excellent agreement with that measured by studying the formation of CN radicals from N_2 in a carbon furnace (Brewer, Templeton and Jenkins 1951). Knowing the heat of formation of the CN radical it is immediately possible to calculate the value of D(NC–CN) from the equation

$$D(\text{NC–CN}) = 2\Delta H_f^0(\text{CN}) - \Delta H_f^0(C_2N_2).$$

In this way there is obtained a value of 4·84 ev for D(NC–CN). Likewise there results a value of 4·98 ev for D(H–CN).

3.4.2. *More complex molecules.*

Most of the bond dissociation energies just discussed were determined by the direct electron impact method, i.e. equation (3) was used directly because the ionization potential of one of the atoms was known. This direct method has only been applied to a few polyatomic molecules because until recently values for the ionization potentials of only a small number of radicals were available. Of these few data only the ionization potential of the methyl was known to be correct to within about 0·1 ev and so until recently only bond dissociation energies of CH_3–X bonds have been capable of being estimated by the direct method. As mentioned earlier, Lossing and his colleagues have recently measured the ionization potentials of several hydrocarbon and other radicals and these data can now be used to determine the dissociation energies of many interesting types of bonds by the direct method. Because these data are available it has now become an urgent necessity that accurate methods be developed for the estimation of the kinetic energy of ions formed in dissociation processes. Only when such complete data are available will the direct method be utilized to the fullest.

Apart from the direct method of estimating bond dissociation energy there is the indirect method introduced and widely used by Stevenson (1942) and Hipple and Stevenson (1943). This method uses thermochemical data in conjunction with electron impact data to derive dissociation energies without knowing the ionization potential of the radical from which the particular ion is formed. Thus if it is desired to determine the dissociation energy $D(R_3–R_4)$ then by determining the appearance potential of the R_1^+ ion from, say, the molecule R_1R_2 and also from R_1R_3, and assuming that this ion is produced by the following process

$$R_1R_2 + e = R_1^+ + R_2 + 2e + A_1 \qquad \ldots\ldots(10)$$

$$R_1R_3 + e = R_1^+ + R_3 + 2e + A_2 \qquad \ldots\ldots(11)$$

and that *no excess kinetic energy* is included in the value of the appearance potential, one can deduce the relation

$$R_1R_2 - R_1R_3 = R_2 - R_3 + A_1 - A_2. \qquad \ldots\ldots(12)$$

If the heats of formation of the compounds R_1R_2, R_3R_4, R_1R_3 and R_2R_4 are known,

and also the dissociation energy $D(R_2–R_4)$ then the value of the heat of the reaction of process (13) can be calculated

$$R_1R_2+R_3R_4=R_1R_3+R_2+R_4+\Delta H. \qquad \ldots\ldots(13)$$

On adding equations (12) and (13) it is easily seen that the process $R_3R_4=R_3+R_4$, i.e. $D(R_3–R_4)$, is given by the equation

$$D(R_3–R_4)=\Delta H+A_1-A_2$$

where $\Delta H=\Delta H_f^0(R_1R_3)+\Delta H_f^0(R_2)+\Delta H_f^0(R_4)-\Delta H_f^0(R_1R_2)-\Delta H_f^0(R_3R_4)$.

Compounds can always be chosen so that R_2R_4 is H_2, CH_4 or some other compound for which there is available a reliable value for $D(R_2–R_4)$. This method has been greatly used by Stevenson to derive values for the dissociation energies of various C–H and C–C bonds in a large number of hydrocarbons. The results which he has obtained are in good agreement with those given by other methods where such estimates are available and this, of course, tends to give one confidence in relying on values deduced by this method when no other estimates are available. The method has the great advantage that so long as only bond dissociation energies are to be determined, then provided the appearance potentials are accurately measured, they need not be determined absolutely for only the difference between the two values is used in the calculation. This method has also been used by Stevenson (1951, 1953) to enable values for the ionization potentials of certain hydrocarbon free radicals to be deduced from measurements of the appearance potentials of the corresponding ions. In this case, however, it is necessary to know the appearance potentials on an absolute voltage scale. As an example of the method we may mention that the combination of the appearance potential of the $C_3H_7^+$ ion (assumed to be *sec*-propyl ion which seems likely), $A(C_3H_7^+)$ in the mass spectra of isobutane, isopentane and 2:3 dimethylbutane with the thermochemical data

$$\text{2:3 dimethylbutane}+CH_4=\textit{iso}\ C_4H_9+C_3H_8,\ \Delta H_{298}^0=0{\cdot}18 \text{ ev}$$

$$\text{2:3 dimethylbutane}+C_2H_6=\textit{iso}\ C_5H_{12}+C_3H_8,\ \Delta H_{298}^0=0{\cdot}40 \text{ ev}$$

and the bond dissociation energies $D(CH_3–H)=4{\cdot}42$ ev, $D(C_2H_5–H)=4{\cdot}20$ ev, leads to the results that $D(\textit{sec}\ C_3H_7–H)=4{\cdot}0_7\pm0{\cdot}1$ ev and $I(\textit{sec}\ C_3H_7)=7{\cdot}45\pm0{\cdot}1$ v.

The above methods have enabled estimates of a large number of bond dissociation energies to be made and we shall now discuss some of these. It will not be possible to consider each case in detail and we shall only discuss those cases which present some interesting features or which are of themselves intrinsically important.

Methane. From a physico-chemical standpoint it is very important to know the energy required to remove a hydrogen atom from methane, i.e. $D(CH_3–H)$. Electron impact data enable this to be estimated by either the indirect or the direct methods for the necessary thermochemical data are available and the ionization potential of the methyl radical is also accurately known. The value found by the direct method (McDowell and Warren 1951, Lossing, Ingold and Henderson 1954a, b) is $102{\cdot}8\pm0{\cdot}9$ kcal mole^{-1}, and the average value given by the application of the indirect method to five different processes (Stevenson 1951) is $101{\cdot}9\pm0{\cdot}9$ kcal mole^{-1}. Both of these electron impact values are in excellent agreement and they also agree well with the value of $102{\cdot}1\pm1{\cdot}0$ kcal mole^{-1} obtained by Kistiakowsky and Van Artsdalen (1944) from a detailed study of the kinetics of the bromination

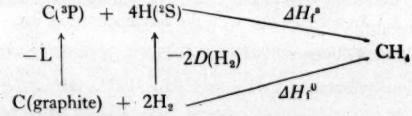
of methane. It is also to be noted that this agreement between the various estimates for $D(CH_3\text{–}H)$ means as a corollary that in the dissociation of methane to form CH_3^+ ions as the result of electron impact, the ion is produced without excess kinetic energy.

The following three bond dissociation energies, $D(CH_2\text{–}H)$, $D(CH\text{–}H)$ and $D(C\text{–}H)$, are of great interest and estimates of these have been made from electron impact studies on methane (McDowell and Warren 1951) and the methyl radical (Waldron 1954). Besides being intrinsically interesting, these three dissociation energies and $D(CH_3\text{–}H)$ are of importance in that if they were known accurately it would then be possible to obtain an indirect estimate of the latent heat of vaporization of carbon. There has been considerable controversy as to the value of this quantity and the problem is as yet not completely solved. Much has been written on this topic (Herzberg 1937, Hagstrum 1951, Long and Norrish 1946, Gaydon 1952, Long 1949) and we shall not attempt to discuss it in any detail. Suffice it to say that most of the indirect evidence seems to point to a value of about 135–145 kcal mole^{-1} and the direct measurements lead to either the value of 141 kcal mole^{-1} (Doehaerd, Goldfinger and Waelbroeck 1952) or about 170 kcal mole^{-1}.

The relationship between the four dissociation energies $D(CH_3\text{–}H)$, $D(CH_2\text{–}H)$, $D(CH\text{–}H)$ and $D(C\text{–}H)$ and the latent heat of vaporization of carbon is obtained as follows. The heat of formation of methane from the gaseous atoms $C(^3P)$ and $H(^2S)$ in their standard thermochemical states is

$$C(g)+4H=CH_4,\ \Delta H_f^a.$$

The normal heat of formation of methane from the elements in their standard thermochemical states is ΔH_f^0. ΔH_f^0 and ΔH_f^a are related by the cycle



i.e. $\Delta H_f^a = L + 2D(H_2) + \Delta H_f^0$ where L is the latent heat of vaporization of carbon and $D(H_2)$ is the dissociation energy of the hydrogen molecule. The sum of the four dissociation energies now under discussion are related to ΔH_f^a by the equation

$$D(CH_3\text{–}H)+D(CH_2\text{–}H)+D(CH\text{–}H)+D(C\text{–}H)=\Delta H_f^a.$$

Thus if all these four dissociation energies were available L could be calculated for $\Delta H_f^0(CH_4)$ and $D(H_2)$ are both accurately known.

As we have mentioned above $D(CH_3\text{–}H)$ is accurately known, and the values for $D(CH_3\text{–}H)$ of $3{\cdot}45 \pm 0{\cdot}2$ ev (McDowell and Warren 1951) and $3{\cdot}93 \pm 0{\cdot}22$ ev (Waldron 1954) are in good agreement. $D(C\text{–}H)$ is known to be 3·6 ev from spectroscopic work (Herzberg 1950). There is, however, some doubt about the value of $D(CH\text{–}H)$. McDowell and Warren (1951) obtained a value of $3{\cdot}4 \pm 0{\cdot}3$ ev on the assumption that the CH^+ ion in the mass spectrum of methane arises by the process:

$$CH_4+e \rightarrow CH^+ + 3H + 2e.$$

Waldron (1954) obtains the value of $5{\cdot}6 \pm 0{\cdot}4$ ev for $D(CH\text{–}H)$ on the assumption

that the CH^+ ion in the mass spectrum of the methyl radical is formed by the process:

$$CH_3+e \rightarrow CH^+ + H_2 + 2e.$$

If the lower of these two estimates is accepted, there results a value of $L \sim 125$ kcal mole^{-1}, the higher value leads to a value of $L \sim 170$ kcal mole^{-1}. A value of $L = 140$ kcal mole^{-1} requires D(CH–H) to be about 3·91 ev. The available electron impact data do not enable a decision to be made between the values for L, but more detailed studies of the electron induced dissociation of CH_4 and the CH_3 and CH_2 radical should certainly help in deciding between the various values.

Higher hydrocarbons. By combining appearance potentials with the relevant thermochemical data, Stevenson (1942, 1951, 1953) has obtained estimates for the dissociation energies of various carbon–hydrogen bonds in several of the higher hydrocarbons. The values which he deduced are in good agreement with those obtained from other sources. Some of Stevenson's values are given in table 5.

Table 5. Dissociation Energies of Carbon–Hydrogen Bonds in Hydrocarbons

Bond	Dissociation Energy (ev)
CH_3–H	4·42±0·04
C_2H_5–H	4·20±0·04
n–C_3H_7–H	4·30±0·09
n–C_4H_9–H	4·38±0·09
t–C_4H_9–H	3·88±0·1

From the same experimental data it is also possible to calculate values for the ionization potentials of free radicals by the indirect method discussed above. The results in some of the cases agree with the directly determined ionization potentials of the radicals. Further details are given in the papers of Stevenson (1942, 1951, 1953) and Lossing and his collaborators (1954).

Alkyl halides. Though the indirect method has been used most successfully by Stevenson to estimate values for the dissociation energies of various bonds in hydrocarbon molecules, this has only been possible because extremely reliable values are available for the heats of formation of hydrocarbons. If accurate thermochemical data are not available the indirect method cannot be used. Unfortunately few accurate values of the heats of formation of the alkyl halides are available and as reliable values were known for the ionization potentials of only the methyl and ethyl radicals little use could be made of the direct method of evaluating bond dissociation energies from electron impact data. The recent determination by Lossing and his collaborators (1954) of values for the ionization potentials of several hydrocarbon radicals has enabled the dissociation energies of several carbon–halogen bonds to be estimated. Several of these new values are given in table 6.

3.5. *Collision Processes involving Negative Ions*

Collision processes in water vapour and ammonia have been studied by Mann, Hustrulid and Tate (1940) following earlier work, to which these authors refer. The results are given in tables 7 and 8.

In the case of water vapour, which is important in many discharge problems, there are several features of interest. The O^+ curve (figure 16) shows a critical

potential of $18{\cdot}8 \pm 0{\cdot}5$ ev, due to $H_2O \rightarrow H_2 + O^+$ (a minimum energy of 18·6 ev is required) and a break at $28{\cdot}1 \pm 1$ ev due, it is suggested to $H_2O \rightarrow 2H + O^+$ (there are excited states of O^+, $^2D^0$ and $^2P^0$, corresponding to 3·3 and 5·0 ev respectively). The H^- curve (figure 17), which (table 7) only shows an abundance at the maximum

Table 6. Dissociation Energies of some Carbon–Halogen Bonds

Compound	Bond Dissociation Energy (ev)	Authors
Methyl iodide	$2{\cdot}20 \pm 0{\cdot}7$	Lossing, Ingold and Henderson (1954b)
,, ,,	$2{\cdot}28 \pm 0{\cdot}15$	McDowell and Cox (1952). Branson and Smith (1953)
	$2{\cdot}33 \pm 0{\cdot}2$	
Methyl bromide	$2{\cdot}88 \pm 0{\cdot}1$	Lossing, Ingold and Henderson (1954b). Branson and Smith (1953)
Methyl chloride	$3{\cdot}49 \pm 0{\cdot}15$	Lossing, Ingold and Henderson (1954b). Branson and Smith (1953)
Methyl fluoride	$4{\cdot}64 \pm 0{\cdot}5$	Lossing, Ingold and Henderson (1954b)
Allyl iodide	$1{\cdot}48 \pm 0{\cdot}7$	,,
Allyl bromide	$2{\cdot}07 \pm 0{\cdot}12$	,,
Allyl chloride	$2{\cdot}62 \pm 0{\cdot}1$	,,
Benzyl iodide	$1{\cdot}50 \pm 0{\cdot}15$	,,
Benzyl bromide	$1{\cdot}94 \pm 0{\cdot}15$	,,
Benzyl chloride	$2{\cdot}62 \pm 0{\cdot}2$	,,
sec-Propyl iodide	$1{\cdot}84 \pm 0{\cdot}1$	Farmer and Lossing (1954, unpublished)
sec-Propyl bromide	$2{\cdot}55 \pm 0{\cdot}1$	,,
sec-Propyl chloride	$3{\cdot}18 \pm 0{\cdot}1$	,,
Propargyl iodide	$1{\cdot}98 \pm 0{\cdot}15$	,,
Propargyl bromide	$2{\cdot}51 \pm 0{\cdot}15$	,,

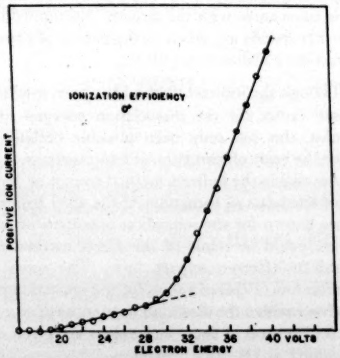


Figure 16. Portion of the ionization efficiency curve of O^+ from H_2O. (Reproduced from *Phys. Rev.*, **58**, 340.)

of the peak of less than 1% of H_2O^+, shows a resonance capture process, having two peaks. H^- from $H_2O \rightarrow OH + H^-$ should appear at 5·7 ev (using kinetic energy data of Lozier 1930); the present value is $5{\cdot}6 \pm 0{\cdot}5$ ev. The second peak in figure 17, i.e. at about 9 ev, has not yet been identified with certainty. The O^- curve from

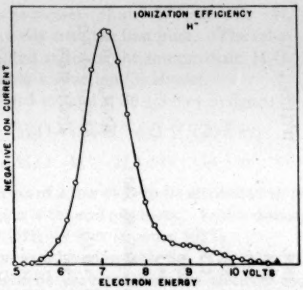


Figure 17. Ionization efficiency curve for H^- from H_2O. (Reproduced from *Phys. Rev.*, **58,** 340.)

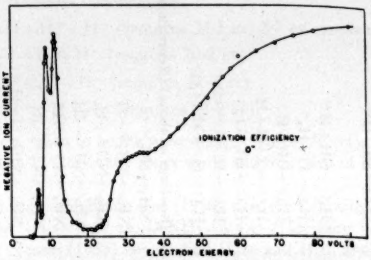


Figure 18. Ionization efficiency curve for O^- from H_2O. (Reproduced from *Phys. Rev.*, **58,** 340.)

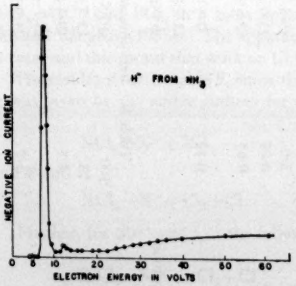


Figure 19. Ionization efficiency curve for H^- from NH_3. (Reproduced from *Phys. Rev.*, **58,** 340.)

Table 7. Ions observed in Water Vapour. Electron Affinity of Oxygen taken as 2·2; of Hydrogen as 0·7 volt. (Reproduced from *Phys. Rev.*, **58**, 340.)

Ion	Intensity at 100 v	Appearance Potential (v)	Probable Process	Calculated Minimum Energy (ev)	Excess Energy or Deduced I.P. (ev)
H_2O^+	100	13·0±0·2	$H_2O \rightarrow H_2O^+$		$I(H_2O)=13·0$
OH^+	23·2	18·7±0·2	$H_2O \rightarrow H+OH^+$	$I(OH)+$ 4·91	$I(OH) \leqq 13·8$
O^+	2·0	18·8±0·5	$H_2O \rightarrow H_2+O^+$	18·6	
		28·1±1·0	$H_2O \rightarrow 2H+O^+$	23·1	$W(O^+) \leqq$ 5·0
H_3O^+		13·8±0·5			
H^+	5·0	19·5±0·2	$H_2O \rightarrow OH+H^+$	18·4	$W(H^+) \leqq$ 1·1
H_2^+	0·07	23·0±2·0	$H_2O \rightarrow O+H_2^+$	20·6	
H^-	0·6†	5·6±0·5	$H_2O \rightarrow OH+H^-$	4·7	
			$H_2O \rightarrow O+H+H^-$	8·8	
O^-	0·15	7·5±0·3	$H_2O \rightarrow 2H+O^-$	7·3	
		23·7±0·5	$H_2O \rightarrow H+H^++O^-$	20·8	
† At maximum.		36·0±3	$H_2O \rightarrow H^++H^++O^-$	34·3	

Table 8. Ions observed in Ammonia (headings as above) (Reproduced from *Phys. Rev.*, **58**, 340.)

NH_3^+	100	10·5±0·1	$NH_3 \rightarrow NH_3^+$		$I(NH_3)=10·5$
NH_2^+	78	15·7±0·1	$NH_3 \rightarrow NH_2^++H$	$I(NH_2)+$ 3·8 (?)	$I(NH_2)=11·9(?)$
NH^+	4·5	19·4±0·5	$NH_3 \rightarrow NH^++H_2$	$I(NH)+$ 3·0	$I(NH) \leqq 16·4$
		23·7±0·5	$NH_3 \rightarrow NH^++2H$	$I(NH)+$ 7·5	
N^+	1·7	24·9±0·5	$NH_3 \rightarrow N^++3H$	25·4	
		28·0±0·5(?)	$NH_3 \rightarrow N^{+*}+3H$	27·3	$W \leqq$ 0·7
H^+	0·5	23·3±0·5	$NH_3 \rightarrow NH+H+H^+$	21·0	
		26·9±0·5	$NH_3 \rightarrow N^*+2H+H^+(?)$	26·7	
H_2^+	0·02	15·5±0·5	$H_2 \rightarrow H_2^+$	15·3	
			$NH_3 \rightarrow NH+H_2^+(?)$	18·3	
NH_3^{2+}	0·02	42 ±3	$NH_3 \rightarrow NH_3^{2+}$		$I(NH_3^+) \leqq 31·5$
H^-	0·15	5·8±0·3	$NH_3 \rightarrow N+H_2+H^-$	5·7	
		23·0±1·0	$NH_3 \rightarrow NH^++H+H^-$	23·2	
NH_2^-	0·3	6·0±0·5	$NH_3 \rightarrow H+NH_2^-$		

H_2O (figure 18) shows processes (3) and (5) (see § 1). $H_2O \rightarrow 2H + O^-$ requires 7·3 ev, agreeing satisfactorily with the first peak. The other peaks, it is suggested, arise from different excited states of the intermediate H_2O^{-*} formed by electron capture (figure 1 and table 1, reaction (3) above).

The breaks at $23{\cdot}7 \pm 0{\cdot}5$ ev, and at $36 \pm 3{\cdot}0$ ev in figure 18 apparently are due to

$$H_2O \rightarrow H + H^+ + O^- \quad (20{\cdot}8 \text{ ev})$$

$$\text{and } H_2O \rightarrow H^+ + H^+ + O^- \quad (34{\cdot}3 \text{ ev})$$

and in each case the excess of a few ev is to be attributed to residual kinetic energy and, possibly, to excitation of the end products. In this discussion, $D(O_2) = 5{\cdot}09$ ev, $D(H_2) = 4{\cdot}46$ ev and $D(OH) = 4{\cdot}4$ ev (Sponer 1935).

The ammonia results are shown in table 8. NH_2^- occurs, but not abundantly; the NH_2 molecule might be expected from its electron configuration to behave similarly to fluorine. The H^- curve is shown in figure 19. The capture process, with onset at 5·8 ev is presumably $NH_3 \rightarrow N + H_2 + H^-$ which requires 5·7 ev (EA(H) taken as 0·7 ev). The ion pair mechanism (reaction 5 above) occurs at 23 ev (figure 19). Possible reactions are:

$NH_3 \rightarrow NH + H^+ + H^-$ (requires 20·3 ev, 2·7 ev excess energy).

$NH_3 \rightarrow N^+ + H_2 + H^-$ (requires 20·2 ev).

$NH_3 \rightarrow N^+ + 2H + H^-$ (requires 24·7 ev).

$NH_3 \rightarrow NH^+ + H + H^-$ (requires 23·2 ev).

The last reaction, taken to be the most likely, involves $I(NH)$, which is deduced to be not greater than 17·2 ev in the above work, and from that of Kusch *et al.* (1937) on HCN.

In this work on ammonia, Mann *et al.* (1940), used the following: $D(H_2) = 4{\cdot}46$ ev, $D(N_2) = 7{\cdot}35$ ev, $Q(NH_3) = 0{\cdot}47$ ev, $D(NH) = 3{\cdot}4$ ev (Herzberg 1939).

More recently Neuert (1953) has studied H_2Se and H_2S, and has given values for the electron affinities of S and Se. The paper gives a full discussion of negative ion processes in these gases and provides an interesting comparison with the earlier work on H_2O mentioned above.

Some of the earliest work on complex polyhalides is that of Kusch, Hustrulid and Tate (1937) on $SbCl_3$, $AsCl_3$ and PCl_3 in a mass spectrometer of the type described by Tate, Smith and Vaughan (1935). The apparatus could not be baked out between the various runs, and this meant that work on Cl^+ could not be carried out for $AsCl_3$ and PCl_3. The results, given in table 9, show that XCl_2^+ is produced abundantly. Evidence was given by the above authors for the reactions (leading to X^+)

$$XCl_3 \rightarrow X^+ + 3Cl$$

in the case of Sb, As and P and

$$XCl_3 \rightarrow X^+ + Cl_2 + Cl$$

for the case of Sb only. Further, for Sb^{2+} and Cl^+ the following were respectively suggested:

$$SbCl_3 \rightarrow Sb^{2+} + Cl_2 + Cl$$

and $$SbCl_3 \rightarrow Sb + 2Cl + Cl^+$$

and $$\rightarrow Sb + Cl_2 + Cl^+$$

Vought's (1947) study of $SiCl_4$ was until very recently the most complete study of any complex halide since it includes measurements on both positive and negative ions. Data on the ions found with a mass spectrometer and a suggested energy level arrangement are given in table 10 and figure 20 respectively.

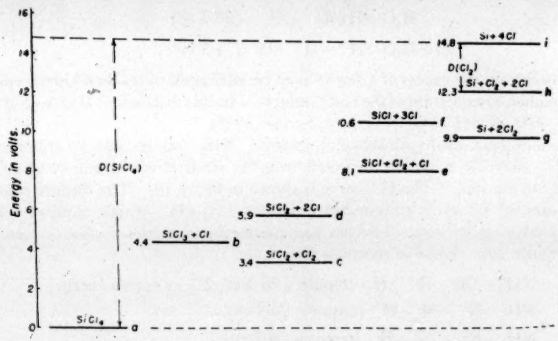


Figure 20. Energies of the possible configurations of one silicon and four chlorine atoms relative to that of the parent molecule, $SiCl_4$. (Reproduced from *Phys. Rev.*, **71**, 93.)

In the interpretation of the data, Vought used the value Q ($SiCl_4$), i.e. the heat of formation of $SiCl_4$, of 142·5 kcal mole^{-1} from Bichowsky and Rossini (1936). This latter work is of great value in work on polyatomic molecules although the need for more thermochemical data is obvious and pressing. Then, since

$$D(SiCl_4) = Q(SiCl_4) + 2D(Cl_2) + L(Si)$$

where L(Si) is the heat of sublimation of Si, we have $D(SiCl_4) = 14{\cdot}8$ ev. (L(Si) = 85 kcal mole^{-1}). Thus levels i, h, g are fixed (figure 20). D(SiCl) was taken as 4·2 ev (Herzberg 1939, Jevons 1936) so that levels f and e are also fixed. The setting of levels c and d which involves $D(SiCl_2)$ is not so satisfactory. Vought, using a relation which applies to the tin chlorides, put $D(SiCl_2) = \frac{3}{5} D(SiCl_4)$,

Table 9. Ions formed in $SbCl_3$, $AsCl_3$ and PCl_3 by Electron Impact. i = Relative intensity at an electron energy of 120 v. A = Appearance potential (v). (Reproduced from *Phys. Rev.*, **52**, 840.)

	Sb		As		P	
Ions	i	A	i	A	i	A
XCl_3^+	100	11·4	100	12·3	100	12·2
XCl_2^+	171	12·3	166	13·0	182	12·5
XCl^+	30	16·4	26·4	17·0	38	17·5
X^+	40	17·0, 19·8	17·2	21·6	36	22·1
XCl_2^{2+}	5·2	32·1	3·8	33·1	2·4	
XCl^{2+}	4·5	34·8	0·6		0·4	
X^{2+}	0·4	38·0				
Cl^+	16·7	20·7, 24·0				

Table 10. Ions formed in $SiCl_4$ (Reproduced from *Phys. Rev.*, **71**, 93.)

Ion	Relative Int. at 75 v	Appearance Potential (v)	Suggested Process: $SiCl_4 \rightarrow$	Min. Energy reqd. (v)	Excess Energy or Deduced I. P. (v)
$SiCl_4^+$	56	$11{\cdot}6 \pm 0{\cdot}2$	$SiCl_4^+$	$I(SiCl_4)$	$I(SiCl_4) = 11{\cdot}6$
$SiCl_3^+$	100	$12{\cdot}9 \pm 0{\cdot}2$	$SiCl_3^+ + Cl$	$4{\cdot}4 + I(SiCl_3)$	$I(SiCl_3) \leqslant 8{\cdot}5$
			$SiCl_3^+ + Cl^-$	$0{\cdot}6 + I(SiCl_3)$	$I(SiCl_3) \leqslant 12{\cdot}3$
$SiCl_2^+$	4·1	$11{\cdot}8 \pm 0{\cdot}3$	$SiCl_2 \rightarrow SiCl_2^+$	$I(SiCl_2)$	$I(SiCl_2) = 11{\cdot}8$
		$18{\cdot}4 \pm 0{\cdot}3$	$SiCl_2^+ + Cl_2$	15·2	$K+E = 3{\cdot}2$
			$SiCl_2^+ + 2Cl$	17·7	$K+E = 0{\cdot}7$
$SiCl^+$	13	$20{\cdot}5 \pm 0{\cdot}3$	$SiCl^+ + Cl_2 + Cl$	$8{\cdot}1 + I(SiCl)$	$I(SiCl) \leqslant 12{\cdot}4$
			$SiCl^+ 3Cl$	$10{\cdot}6 + I(SiCl)$	$I(SiCl) \leqslant 9{\cdot}9$
Cl^+	5·5				
Si^+	14	27 ± 1	$Si^+ + 2Cl_2$	18·0	$K+E = 9{\cdot}0$
			$Si^+ + Cl_2 + 2Cl$	20·4	$K+E = 6{\cdot}6$
			$Si^+ + 4Cl$	22·9	$K+E = 4{\cdot}1$
$SiCl_2^-$	0·25	$0{\cdot}8 \pm 0{\cdot}3$	$SiCl_2^- + Cl_2$	5·9	$EA(SiCl_2) \geqslant 5{\cdot}1$
			$SiCl_2^- + 2Cl$	3·4	$EA(SiCl_2) \geqslant 2{\cdot}6$
			$SiCl_2 \rightarrow SiCl_2^-$		
Cl^-	4·0	$1{\cdot}4 \pm 0{\cdot}3$	$Cl^- + SiCl_3$	0·6	$K+E = 0{\cdot}8$
		$2{\cdot}4 \pm 0{\cdot}3$	$Cl^- + SiCl_2 + Cl$	2·1	$K+E = 0{\cdot}3$
		$8{\cdot}2 \pm 0{\cdot}3$	$Cl^- + SiCl + Cl_2$	4·3	$K+E = 3{\cdot}9$
			$Cl^- + SiCl + 2Cl$	6·8	$K+E = 1{\cdot}4$
$SiCl_3^{2+}$	4·9	$33{\cdot}8 \pm 0{\cdot}5$			
$SiCl_2^{2+}$	0·65	$37{\cdot}3 \pm 0{\cdot}5$			
$SiCl^{2+}$	1·2	42 ± 1			

following Parti and Samuel (1937). Finally, from the work of Asundi, Karim and Samuel (1938), $D(SiCl_3)$ may be estimated from the Si–Cl bond energy which is taken as 4·4 ev.

The important reaction (table 10) involving $SiCl_3^+$ is of interest, and Vought's discussion of it illustrates the difficulties that often arise in work of this kind. Since Cl^- and $SiCl_3^+$ are both observed to appear at about 13 ev, the reaction

$$SiCl_4+e \rightarrow SiCl_3^+ + Cl^- + e$$

is assumed to occur, as in the case of other molecules (such as CCl_4 etc.) discussed in this Report. Since rough measurements of ion currents show that the $SiCl_3^+$ current is much greater than that of Cl^- just above 13 ev, it is assumed that

$$SiCl_4+e \rightarrow SiCl_3^+ + Cl + 2e$$

also occurs, even though no break in the $SiCl_3^+$ current curve is noticed at 3·8 ev above the critical (appearance) potential for $SiCl_3^+$. It would, therefore, be necessary either to obtain more accurate abundance measurements or to devise, for example, a method of detecting free Cl atoms before the relative importance of these reactions can be assessed.

The production of Si^+ at 27 ev is most probably by $SiCl_4 \rightarrow Si^+ + 4Cl$ since the other suggested mechanisms (table 10) involve still more kinetic energy. Indeed, Vought points out, all processes except that giving $SiCl^+$ are most easily explained by processes giving the maximum number of products. For example the Cl^- peak (figure 21) at 6 ev is explained better by $SiCl_4+e \rightarrow SiCl+2Cl+Cl^-$ rather than by $SiCl_4+e \rightarrow SiCl+Cl_2+Cl^-$.

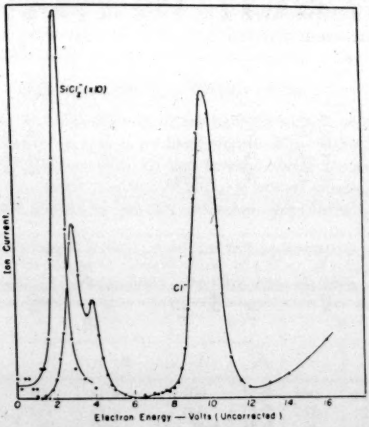


Figure 21. Appearance potential curves for negative ions in $SiCl_4$. (Reproduced from *Phys. Rev.*, **71**, 93.)

The formation of $SiCl_4^+$ is of interest, see table 11, when compared with CCl_4^+ from CCl_4 (Baker and Tate 1938).

Table 11. Comparative Data for certain Products of Ionization in CCl_4 and $SiCl_4$

Dissociation products	Appearance potential (v)	Relative intensity at 75 ev
CCl_4^+	11 $\pm$1·0	0·02
$CCl_3^+ + Cl^-$	12·2$\pm$0·2	100
$SiCl_4^+$	11·6$\pm$0·2	56
$SiCl_3^+ + Cl^-$	12·9$\pm$0·2	4–100†

† This range of intensities is given by Vought because of the uncertain collection efficiency of his mass spectrometer.

Vought suggests that the relatively probable formation of $SiCl_4^+$ is because the bonding in $SiCl_4$ is partially double bond and partially ionic (Pauling 1940) and so presumably stronger than in CCl_4, where there is valency bonding. Other qualitative arguments of a similar character may be involved to explain the relatively easy removal of an electron from $SiCl_4$ but these would be more convincing if quantitative estimates of cross sections could be made.

We propose next to describe the experimental work on organic halides and related molecules which are of interest for various reasons. The study of methyl and ethyl halides is of basic importance in relation to the strength of carbon bonds and the structure of various saturated hydrocarbons and their derivatives. Also, the electrical properties of gases such as SF_6, CCl_2F_2 (freon 12) and CCl_4 are of considerable technical interest (Meek and Craggs 1953), because of their use in Van de Graaff high voltage d.c. generators and in gas-filled cables. Further, since experimental and theoretical studies of negative ions in gaseous discharges or in fundamental collision processes, and possibly also in the electrical characteristics of liquids, have scarcely advanced beyond the initial semi-qualitative stage, these experiments on polyatomic molecules may be of interest. It must be emphasized here that it is clearly necessary first to pursue work on the simplest molecules (notably O_2 and the halogens) before the properties of more complex molecules can have much chance of elucidation. It is notable that no precise theoretical values of electron attachment cross sections exist even for O_2, and the state of the experimental work is little better. Further, direct (Bradbury 1933) and indirect (Harrison and Geballe 1953) methods for the measurement of attachment coefficients in swarm conditions have so far not given satisfactory agreement. It is hoped that work now being carried out on the measurement of cross sections with a Lozier apparatus (see § 2) at Liverpool will give further data for comparison with those of the latter authors but the position is unsatisfactory as Massey has previously pointed out (1950) and there is great scope for experimental and theoretical workers. Despite this tenuous body of evidence for even the simple diatomic molecules we shall show that some complex molecules behave, in their most important ionization processes, effectively as diatomic molecules although there are many important exceptions.

Direct measurements of ionization cross sections in halide molecules and in polyatomic molecules which form negative ions in collision processes (e.g. NH_3 and C_2N_2) appear, surprisingly, to be limited to preliminary work on UF_6 and UCl_4 (Berkey *et al.* 1949). The experiments were carried out during 1944–45 for special reasons and thus are not final. The technical difficulties are also exceptionally great, but an ionization cross section of about 5×10^{-15} cm^2 at 100 ev (electron impact) was obtained for UCl_4, a value about ten times that for argon in comparable

conditions. The maximum ionization cross section for formation of Hg^+ by electron impact is (Massey and Burhop 1952, p. 38) about 6×10^{-16} cm^2.

We therefore pass to a review of the work on the identification of the various collision processes in electronegative polyatomic molecules. Apparently the first measurements in this field were on CO_2, N_2O and NO_2 by Kallmann and Rosen (1930) and Smyth and Stueckelberg (1930). A comprehensive list of references to work of this kind is given by Massey and Burhop (1952, Appendix 2). The first study of a halide molecule appears to be that of Bauer and Hogness (1935) for CH_3Cl. Baker and Tate (1938) worked with CCl_4 and CCl_2F_2 and their data for CCl_4 are given as an example, together with that of Warren and Craggs (1952) in table 12.

Table 12. Collision Process Data for CCl_4

	Relative abundance at 75 ev		Appearance potential (ev)	
Ion	B. & T.	W.&C.	B. & T.	W. & C.
CCl_4^+	0·02	—	11·0±1	—
CCl_3^+	100	100	12·2±0·2	11·8±0·05
CCl_2^+	23·5	20	16·0±0·2	16·1±0·1
CCl^+	27·8	22	17·1±0·2 19·5±0·2	19·3±0·2
Cl_2^+	0·35	1·1	16·4±0·5 23·0±1	—
Cl^+	27·6	14	19·1±0·2	20·8±0·3
C^+	14·4	2·8	23·5±0·2	26·6±0·3
CCl_3^{2+}	1·6	3·8	31·8±1	
CCl_2^{2+}	0·8	—	34 ±1	—
Cl^+	99·5		1·3±0·5 5·8±0·5 12·4±0·2	—
Cl_2^+	0·1		4·3±1 17·0±0·5	—
C^+	0·01		—	
CCl^+	0·005		—	

B. & T. = Baker and Tate; W. & C. = Warren and Craggs.

It will be noted that Cl^- predominates amongst the negative ions and CCl_3^+ amongst the positive ions. Thus with an electron beam of 12 ev energy the only ions of any importance (CCl_4^+ is practically unobservable) are CCl_3^+ and Cl^-. Similar behaviour is noticed for the whole of the series CCl_4, CCl_3F, CCl_2F_2, $CClF_3$ and CF_4 which have been studied in Liverpool (Warren and Craggs 1952). The primary positive ion is in every case negligible, i.e. less than 1 in 1,000 of the most abundant positive ion, except for $CClF_3^+$ from $CClF_3$ when the relative abundance, cf. CF_3^+, is about 0·2%; the preponderance of the most probable ion is demonstrated by figure 22 for CCl_3F. Exceptions to this rule are $SiCl_4$ (Vought 1947, Marriott 1954) UCl_4 and more complex molecules such as UF_6 (Burhop *et al.* 1949) and SF_6 (Ahearn and Hannay 1953, Marriott 1954) and $GeCl_4$ (Marriott 1954). Similarly, the negative ion spectrum is generally simple as shown in table 11, above, although again there are important exceptions such as SF_6 and UF_6 where complex negative ions such as SF_6^- and SF_5^- and UF_n^- are

formed in greater quantities than the simplest negative ion F^-. UCl_4 is especially complicated (Burhop *et al.* 1949) and UCl_4^-, UCl_3^-, UCl_2^- and Cl^- are all formed.

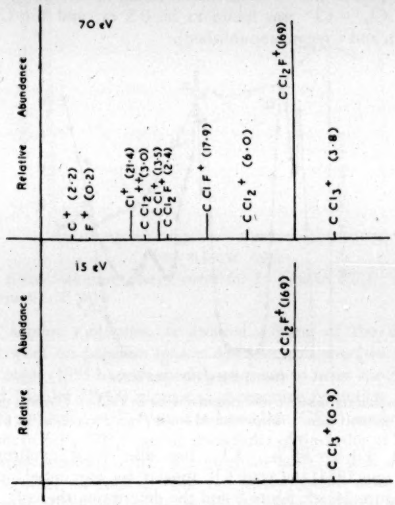


Figure 22. Positive ion mass spectra for CCl_3F. Relative abundances in arbitrary units are in parenthesis. (Reproduced from *Mass Spectrometry* (Institute of Petroleum).)

Collected data on the primary positive ions from certain tetrachlorides are given in table 13, where the strength of the most abundant ion XCl_3^+ is taken as 100. The electron energy is 75 ev.

Table 13. Abundance Ratios of certain Positive Ions in various Tetrachlorides

	CCl_4(*a*)	$SiCl_4$(*a*)	$GeCl_4$(*a*)	UCl_4(*b*)	$TiCl_4$(*c*)
XCl_4^+ as % of XCl_3^+	0	64	35	13·5	58

(*a*) Marriott 1954, (*b*) Burhop *et al.* 1949 and (*c*) Thorburn, unpublished, at 70 ev.

The other halogen derivatives of methane present results generally similar to those for CCl_4, and details have been given for CCl_2F_2 by Baker and Tate (1938) who determined the heat of formation to be 108 ± 23 kcal mole^{-1} as compared with the very recent figure of 112 kcal mole^{-1} found thermochemically by Kirkbride and Davidson (1954). This latter publication gives valuable data for the heats of formation of CF_4, CF_3Cl, CF_2Cl_2, $CFCl_3$ and CCl_4. Further data of this kind for other similar molecules would be invaluable as we shall show later in a few specific examples.

The principal processes involving Cl^- in CCl_4 have been investigated with a Lozier apparatus (Marriott, Thorburn and Craggs 1954) with results shown in figure

23. The appearance of Cl^- in the resonance capture process is at $0{\cdot}3 \pm 0{\cdot}2$ ev if allowance for the initial kinetic energy is made. The peaks at higher energies are presumably due to processes in which other Cl atoms are removed with the Cl^- ion from the CCl_4 molecule. The minimum initial kinetic energy for Cl^- in the process $CCl_4 \rightarrow CCl_3^+ + Cl^-$ was found to be 0·5 ev, and for CCl_3^+ less than 0·4 ev (Thorburn and Craggs, unpublished).

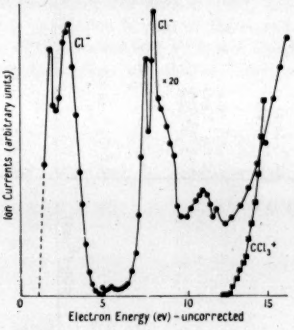


Figure 23. Predominating positive and negative ions in CCl_4. Current units are different for ions of different sign. (Reproduced from *Proc. Phys. Soc.* B, **67**, 437.)

The reaction $TiCl_4 \rightarrow TiCl_3^+ + Cl^-$ has also been confirmed (Marriott, Thorburn and Craggs 1954) and again Cl^- appears, for zero initial kinetic energy, at $0 \pm 0{\cdot}1$ ev (see figure 24; cf. figure 8 and the description thereof). This is to be

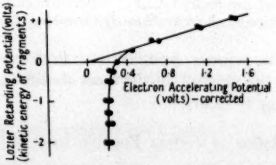


Figure 24. Lozier plot for Cl^- ions in $TiCl_4$. (Reproduced from *Proc. Phys. Soc.* B, **67**, 437.)

expected if the reaction products are unexcited and if the electron affinity of Cl is greater than the $TiCl_3$–Cl bond dissociation energy. Further evidence of the formation of negative ions at electron energies less than 1 ev is provided by the mass spectrometer data accumulated in this laboratory. A striking example is the formation of I^- from CF_3I (Marriott and Craggs 1953) shown in figure 25; Br^- from CF_3Br exhibits a very large low energy peak also (figure 26, taken from the above work) but in this case the formation of negative ions by the ion pair (dissociation) process is negligible (cf. I^- from CF_3I) although (figure 27) F^- from CF_3Br shows both resonance capture and ion pair processes, together with multiple capture peaks due to dislodgement of successive F atoms. The full and quantitative explanation of these effects, and the striking differences in the negative ion patterns

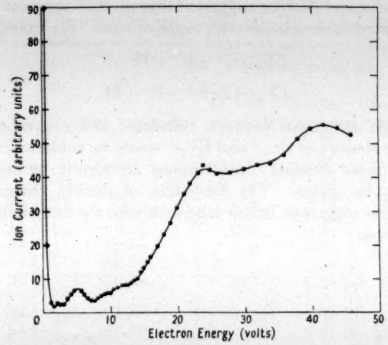


Figure 25. Ion current–electron energy curve for I^- ions in CF_3I. (Reproduced from E.R.A. Report L/T 301.)

for apparently similar molecules, is awaited. Some of the difficulties to be encountered in work on negative ions at low electron energies are discussed by Marriott and Craggs (1954 b) with special reference to mass spectrometer work.

Dibeler and Mohler (1948) measured appearance potentials for positive ions only in SF_6, CF_4 and SiF_4. SF_6^+ was not detectable, CF_4^+ was present to less than 2% of CF_3^+ and, in SiF_4, SiF_4^+ had an abundance of the order of 1 in 10^3 of SiF_3^+. The detailed mass spectra for SF_6 are given in table 14 and similar data were obtained for the other molecules. The full interpretation of the data was again impossible because of the lack of the requisite thermochemical and spectroscopic data, and the kinetic energies of the fragments in the various reactions were not measured. Vought's (1947) data, see table 10, demonstrate the great importance of

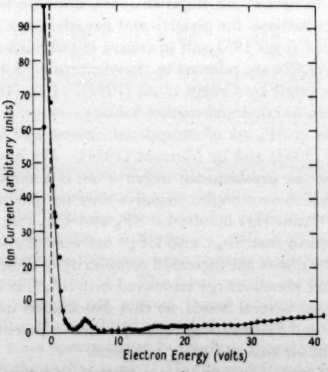


Figure 26. Ion current–electron energy curve for Br^- ions in CF_3Br. (Reproduced from E.R.A. Report L/T 301.)

the latter. Dibeler and Mohler suggested that positive ions, for these molecules, tended to be formed in association with negative ions. For example, the reactions

$$CF_4 \rightarrow C^+ + F^- + 3F$$

and

$$CF_4 \rightarrow C + F^+ + F^- + 2F$$

gave a reasonable agreement between calculated and experimental appearance potentials. The absence of SF_6^+ and CF_6^+ would be consistent with this general mechanism but more detailed considerations are clearly necessary before final conclusions may be drawn. The formation of doubly charged positive ions (SF_4^{2+}, etc.), it was suggested, is also consistent with the formation of negative ions in the same process.

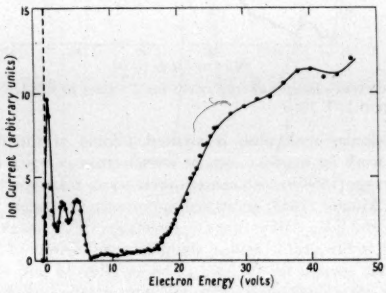


Figure 27. Ion current–electron energy curve for F^- ions in CF_3I. (Reproduced from E.R.A. Report L/T 301.)

BF_3 and BCl_3 have been investigated by Osberghaus (1950) (positive ions) and by Marriott (1954) (positive and negative ions), and the latter author has also obtained dissociation patterns for positive and negative ions in CF_3Br, CF_3I, and $GeCl_4$ (Marriott and Craggs 1953 and in course of publication). The interesting cases of C_2N_2 and CH_3CN are referred to elsewhere (see § 3.4 (*a*)) and the former is also dealt with in detail by Craggs *et al.* (1952). Cox (1953) has investigated negative ion formation in ethyl and methyl halides.

The negative ions in SF_6 are of exceptional interest and have been studied by Ahearn and Hannay (1953) and by Marriott (1954). We have shown above that, for several molecules, the predominant negative ion is a single negatively charged halogen atom and that more complex negative ions are generally negligible at the pressures (10^{-5}–10^{-6} mm Hg) involved. SF_6 and UF_6 (see Burhop *et al.* 1949) prove to be exceptions in that SF_6^- and UF_n^- are very abundant and, it has been suggested, this may be due to the increased complexity of these molecules (cf. CCl_4 for example) so that the excess energy associated with SF_6^-, as in the reaction above, may be spread amongst several bonds, so that dissociation into SF_5 and F^- does not follow. Ahearn and Hannay suggested qualitative potential energy diagrams (see below) to account for this and other reactions.

The SF_6^-, SF_5^- and F^- currents are plotted in figures 28, 30 and 31 for electron energies varying up to about 50 v. Marriott had found that the F^- peak

Table 14. Mass Spectrum of SF_6 for 50, 70, and 100-volt Electrons (Reproduced from *J. Res. Nat. Bur. Stand.*, **40**, 25.)

m/e	Ion	Relative intensities		
		50 volts	70 volts	100 volts
16	S^{2+}	0·03	0·06	0·09
19	F^+	3·37	5·63	9·11
25·5	SF^{2+}	.....	0·76	1·97
32	S^+	5·60	9·96	13·5
33	S^+	0·06	0·09	0·11
34	S^+	0·27	0·47	0·60
35	SF_2^{2+}	1·36	6·80	13·3
35·5	SF_2^{2+}	.....	0·10	0·10
36	SF_2^{2+}	0·07	0·33	0·62
44·5	SF_3^{2+}	0·38	1·26	2·17
45·5	SF_3^{2+}	.....	0·07	0·11
51	SF^+	8·66	11·0	11·6
52	SF^+	0·07	0·09	0·11
53	SF^+	0·40	0·51	0·64
54	SF_4^{2+}	4·05	11·0	17·6
54·5	SF_4^{2+}	.....	0·09	0·11
55	SF_4^{2+}	0·23	0·54	0·78
70	SF_2^+	5·74	6·39	6·46
71	SF_2^+	0·04	0·07	0·09
72	SF_2^+	0·27	0·30	0·27
89	SF_3^+	27·6	33·5	35·7
90	SF_3^+	0·20	0·26	0·24
91	SF_3^+	1·22	1·48	1·56
91·8	$SF_5^+ \rightarrow SF_4^+ + F$	0·21	0·26	0·24
108	SF_4^+	8·98	10·8	11·5
109	SF_4^+	0·07	0·11	0·10
110	SF_4^+	0·41	0·47	0·51
127	SF_5^+	100	113	117
128	SF_5^+	0·92	1·14	1·15
129	SF_5^+	4·30	4·89	4·87
131	SF_5^+	0·04	0·04	0·04

below 2 ev (which is not shown in Ahearn and Hannay's results given in figure 28) is indeed large but still not comparable with the SF_6^- or SF_5^- currents at these low electron energies. The variations of negative ion currents with pressure are shown in figure 29, which includes also a curve for F_2^-, which was formed only in a small quantity. The negative ion currents vary linearly with pressure, except for the SF_6^- and SF_5^- currents at 35 ev. The latter results are mentioned again below. Another interesting negative ion process found to occur was $SF_6^{-*} \rightarrow SF_5^- + F$ which took place after the ions had passed through the accelerating part of the electric field in the mass spectrometer but before appreciable magnetic deflection had taken place (see the earlier work of Hipple and Condon 1945). This process has been mentioned in § 3.1.

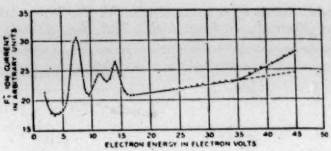


Figure 28. Ion current–electron energy curve for F^- ions in SF_6. (Reproduced from *J. Chem. Phys.*, **21**, 119.)

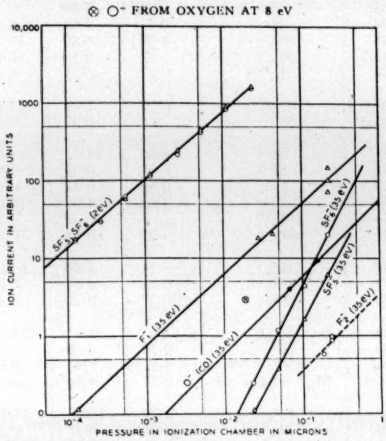


Figure 29. Pressure dependence and relative magnitudes of negative ion currents in SF_6. (Reproduced from *J. Chem. Phys.*, **21**, 119.)

The negative ion processes involved are those given above as reaction (3) and (4) (§ 1). The SF_6^- and SF_5^- peaks at very low electron energies are attributable to process 3(*b*). The possible reactions are

$$e+SF_6 \rightarrow SF_6^{-*} \begin{cases} \nearrow SF_6^- \\ \searrow SF_5^- + F. \end{cases}$$

The branching processes appear to be roughly comparable in strength. Also, it is possible that

$$e+SF_6 \rightarrow SF_6^{-*} \rightarrow SF_5 + F^-$$

$$e+SF_6 \rightarrow SF_6^{-*} \rightarrow SF_4 + F + F^-, \text{ etc.}$$

may take place giving differences in critical potentials of about 4 ev, which is approximately the energy necessary to remove successive F atoms since the dissocia-

tion energy for $SF_6 \rightarrow S+6F$ is about 22·4 ev. The multiple peaks seen in figure 28 are possibly due to this type of reaction since the energy intervals between peaks are about 4 and 3 ev, and the kinetic energy term may provide better agreement when known. On the arguments also, the F^- current should vary linearly with

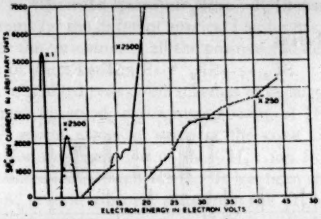


Figure 30. Ion current–electron energy curve for SF_6^- ions in SF_6. (Reproduced from *J. Chem. Phys.*, **21**, 119.)

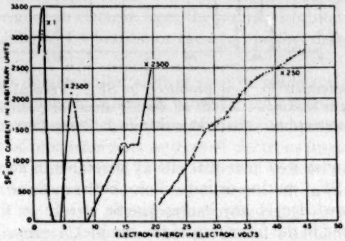


Figure 31. Ion current–electron energy curve for SF_5^- ions in SF_6. (Reproduced from *J. Chem. Phys.*, **21**, 119.)

pressure, in agreement with the data of figure 29. The SF_6^- and SF_5^- currents appearing at about 17·3 ev in figure 30 and 31 are due probably to capture of slow electrons formed in reactions such as

$$SF_6 + e \rightarrow SF_6^* + e \quad \rightarrow SF_5 + F + e, \text{ etc.} \qquad \ldots\ldots(14)$$

$$SF_6 + e \rightarrow SF_6^{+*} + 2e \rightarrow SF_5^+ + F + 2e, \text{ etc.} \qquad \ldots\ldots(15)$$

$$SF_6 + e \rightarrow SF_6^* + e \quad \rightarrow SF_5^+ + F^- + e, \text{ etc.} \qquad \ldots\ldots(16)$$

by SF_6 molecules, in process (3*b*) (§ 1). Ahearn and Hannay point out that this would necessitate a variation with (pressure)^2, as shown in figure 29, and they show also that the shape of the (current, electron energy) curve for SF_5^+ which has an appearance potential of 15·9 ev, is consistent with this mechanism. These correlations are brought out more clearly in figure 32.

The SF_6^- and SF_5^- currents between 2 and 16 ev in figures 30 and 31 can be explained by reaction (14) above, followed by capture of the slow electron so produced. Above 15·9 ev the inflections in the negative ion current curves may be attributable to reactions of the types (15) and (16) above, and should then fit similar inflections on the SF_5^+ curve. The agreement with the Dibeler and Mohler data is satisfactory for SF_3^+.

The appearance of F^- at about 15·7 ev is to be attributed to reaction (16), since SF_5^+ appears at the same electron energy, and the F^- current varies linearly with pressure. The increase in slope (figure 28) at 34·3 ev agrees reasonably well with that of F^+ (Dibeler and Mohler 1948) at 35·8 ev. The suggested reaction is

$$SF_6+e\rightarrow SF_6^*\rightarrow S+4F+F^-+F^+$$

requiring 37·4 ev, whereas

$$SF_6+e\rightarrow SF_6^{+*}\rightarrow S+5F+F^+$$

requires 41 ev.

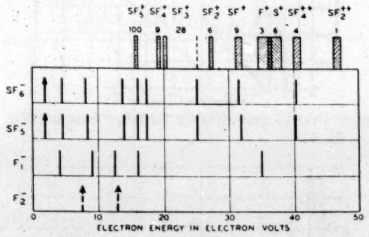


Figure 32. Onset potentials for ions produced in SF_6. Vertical lines represent onsets, arrows represent maxima. Width of shaded areas for positive ion data represents assigned probable errors. (Reproduced from *J. Chem. Phys.*, **21**, 119.)

Ahearn and Hannay give potential energy diagrams to account for the occurrence of SF_6^- and SF_5^- but no definite choice between these may be made at the present. Also, knowledge of any excess kinetic energy in the above work, for example that involved in the formation of the complex negative ions at low electron energies, should be obtained before even tentative conclusions could be reached. Ahearn and Hannay (1953) estimate that, since the ion current of SF_5^- or SF_6^- from SF_6 is about 1,000 times the O^- current formed in CO at 35 ev or about 400 times the O^- current in O_2 at 8 ev, the maximum capture cross section of SF_6 to give SF_6^- is of the order of 10^{-16} cm². This estimate is based on Lozier's (1934) quantitative comparison of O^- and CO^+ currents in CO, and on Tate and Smith's (1932) quantitative data. It seems possible that measurements on Br^- from CF_3Br and on certain other molecules (Thorburn, Marriott and Craggs, unpublished) will, when complete, give capture cross sections of the order of 10^{-15} cm² for very low energy (<1 ev) electrons.

§ 4. Theoretical Studies

Recent molecular orbital calculations have, as has been mentioned earlier, led to considerable interest in the electronic structure of the states of molecular ions. This has fortunately coincided with the development of refined experimental techniques (see § 2) using essentially mono-energetic electrons and so it is evident that there will soon be available reliable data on the structures and energy levels of a large number of interesting molecular ions. Concurrent with these advances there has developed a considerable interest in the exact mode of formation of the various ions observed in the mass spectra of compounds. Thus Laidler (1954) has discussed in detail

the mode of formation of the various ions which are produced when water vapour is bombarded with electrons of known energy. By considering the electronic states of the ions known to be produced, he has been able to apply group theoretical arguments (Mulliken 1933, Shuler 1953) to infer from which states of the H_2O^+ ion the various ions in the mass spectrum arise. By considering the possible potential energy diagrams for the different species, Laidler has been able to suggest acceptable schemes for the formation of all the positive and negative ions observed in the mass spectrum of water vapour.

Similar studies have enabled reasonable schemes to be suggested for the modes of formation of the ions observed in the mass spectra of HCN, NH_3 and CH_4 (McDowell 1953). A recent extensive study of the same kind has enabled the adiabatic correlations for the formation of the CH_3^+ ion from methane and its derivatives to be elucidated (McDowell 1954). It has been shown that the CH_3^+ ion is formed in its $^1A_1$ ground vibronic state from the $^2T_2$ vibronic state of the CH_4^+ ion by a dissociation process occurring at 14·39 ev. Similar studies have also shown that when the CH_3^+ ion is formed from CH_3I at 12·36 ev, it arises from the $^2A_1$ vibronic state of the CH_3I^+ ion. This $^2A_1$ vibronic state of the CH_3I^+ ion results from the removal of an electron from the $[\sigma a_1]$ C–I orbital of methyl iodide, while the $\nu_3(a_1)$ stretching C–I vibration or the $\nu_2(a_1)$ methyl deformation vibration is excited.

Many more similar studies will doubtless soon be made and it is of interest to note that those first mentioned have already proved to have been a powerful stimulus to research in the detection and identification of the excited states of molecular ions.

The development of monoenergetic sources of electrons (mentioned above) has naturally led to a renewal of interest in the shapes of ionization efficiency curves and much work on this topic will doubtless be carried out in the near future. A detailed discussion of the theory of ionization processes is given by Massey and Burhop (1952) in Chapter III of their book. Bates, Fundaminsky, Leech and Massey (1950) have given a lengthy and elegant analysis of the theory of excitation and ionization processes and have shown that the Born approximation leads to a threshold ionization efficiency law which varies with the first power of the excess energy of the ionizing electron. Another, and somewhat different treatment by Wannier (1953) yields the result that the probability of ionization just above the threshold is proportional to $\epsilon^{\frac{1}{2}\mu-\frac{1}{2}}$ where ϵ is the energy and

$$\mu=\tfrac{1}{2}[(100z-9)/(4z-1)]^{\frac{1}{2}}$$

where Z is the charge of the ion formed. For atoms (or molecules) when $Z=1$, this leads to the result that the ionization probability is proportional to $\epsilon^{1\cdot127}$.

There are very few satisfactory experimental data available for the accurate assessment of these theories but recent ionization probability curves for He^+ and Xe^{2+} obtained by the new pulsed technique which yields essentially monoenergetic electrons (Fox *et al.* 1951) have shown that the experimental data obey a linear law rather than a power one (Hickam, Fox and Kjeldaas 1954). These results have, however, only been obtained over a restricted energy range and may not be sufficient to test the Wannier law. Further, the scatter in the experimental results in the region close to the threshold makes it impossible to decide which of the two types of laws gives the best fit to the data. Further careful experimental work is thus highly desirable and studies of the ionization probabilities for molecules are also, of course, necessary.

ACKNOWLEDGMENTS

We are indebted to various colleagues, notably Dr. J. W. Warren, Dr. J. Marriott, Dr. B. C. Cox, Mr. R. Thorburn and Mr. R. Shiels for a fruitful collaboration, extending over a period of more than six years. In this time a good deal of experimental work, partly of an exploratory nature, on complex molecules has been carried out in these and other laboratories and we thank those mentioned above for their efforts in this difficult field.

We are also grateful to Dr. F. P. Lossing (National Research Council of Canada) for his generosity in giving us access to unpublished work, and to the following for permission to reproduce figures and tables: the Electrical Research Association, the Institute of Petroleum and the Editors of the Journal of Chemical Physics, the Journal of Research of the National Bureau of Standards, the Physical Review and the Review of Modern Physics.

REFERENCES

AHEARN, A. J., and HANNAY, N. B., 1953, *J. Chem. Phys.*, **21,** 119.
ARNOT, F. L., 1933, *Collision Processes in Gases* (London: Methuen).
ASUNDI, R. K., KARIM, M., and SAMUEL, R., 1938, *Proc. Phys. Soc.*, **50,** 581.
BAKER, R. F., and TATE, J. T., 1938, *Phys. Rev.*, **53,** 683 (A).
BARNARD, G. P., 1953, *Modern Mass Spectrometry* (London: Institute of Physics).
BATES, D. R., FUNDAMINSKY, A., LEECH, J. W., and MASSEY, H. S. W., 1950, *Phil. Trans. Roy. Soc.*, **243,** 93.
BAUER, S. H., and HOGNESS, T. R., 1935, *J. Chem. Phys.*, **3,** 687.
BERKEY, W. E., BURHOP, E. H. S., CRAGGS, J. D., KEENE, J., and MASSEY, H. S. W., 1949, *The Characteristics of Electrical Discharges in Magnetic Fields* (New York: McGraw-Hill), Chap. 6.
BERRY, C. R., 1950, *Phys. Rev.*, **78,** 597.
BEUTLER, H., and JÜNGER, H. O., 1936, *Z. Phys.*, **100,** 80.
BICHOWSKY, F. R., and ROSSINI, F. D., 1936, *Thermochemistry of Chemical Substances* (New York: Reinhold).
BLEAKNEY, W., 1930, *Phys. Rev.*, **35,** 1180; 1932, *Ibid.*, **40,** 456.
BLEAKNEY, W., CONDON, E. U., and SMITH, L., 1937, *J. Phys. Chem.*, **41,** 197.
BLEAKNEY, W., and SMITH, L., 1936, *J. Phys. Chem.*, **49,** 883.
BRADBURY, N. E., 1933, *Phys. Rev.*, **44,** 883.
BRANSON, H., and SMITH, C., 1953, *J. Amer. Chem. Soc.*, **75,** 4133.
BREWER, L., TEMPLETON, L. K., and JENKINS, F. A., 1951, *J. Amer. Chem. Soc.*, **73,** 1462.
BUCHDAHL, R., 1941, *J. Chem. Phys.*, **9,** 146.
BURHOP, E. H. S., MASSEY, H. S. W., and WATT, C., 1949, *The Characteristics of Electrical Discharges in Magnetic Fields* (New York: McGraw-Hill), p. 145.
CLARKE, E. M., 1954, *Canad. J. Phys.*, **32,** 764.
COX, B. C., 1953, *Thesis*, University of Liverpool.
CRAGGS, J. D., MCDOWELL, C. A., and WARREN, J. W., 1952, *Trans. Faraday Soc.*, **48,** 1093.
CRAIG, R. D., 1954, *Applied Mass Spectrometry*. (London: Institute of Petroleum), p. 230.
DAVIDSON, F. G., and KIRKBRIDE, F. W., 1954, *Nature, Lond.*, **174,** 79.
DIBELER, V. H., and MOHLER, F. L., 1948, *J. Res. Nat. Bur. Stand.*, **40,** 25.
DOEHAERD, P., GOLDFINGER, P., and WAELBROECK, F., 1952, *J. Chem. Phys.*, **20,** 757.
DONNOLLY, B. L., and CARR, H. E., 1954, *Phys. Rev.*, **93,** 111.
DOUGLAS, A. E., 1952, *Canad. J. Phys.*, **30,** 302.
ELLISON, F. O., and SHULL, H., 1953, *J. Chem. Phys.*, **21,** 1420.
ELTENTON, G. C., 1942, *J. Chem. Phys.*, **10,** 403; 1947, *Ibid.*, **15,** 455.
EVANS, A. G., 1946, *Trans. Faraday Soc.*, **42,** 719.
EVANS, M. G., 1952, *Mass Spectrometry* (London: Institute of Petroleum), p. 4.
FARMER, J. B., HENDERSON, I. H. S., MCDOWELL, C. A., and LOSSING, F. P., 1954, *J. Chem. Phys.*, **22,** 1948.

Fox, R. E., Hickam, W. M., and Kjeldaas, T., 1953, *Phys. Rev.*, **69,** 347.
Fox, R. E., Hickam, W. M., Kjeldaas, T., and Grove, D. J., 1951, *Phys. Rev.*, **84,** 859.
Fox, R. E., and Hipple, J. A., 1948, *Rev. Sci. Instrum.*, **17,** 462.
Gaydon, A. G., 1952, *Dissociation Energies and Spectra of Diatomic Molecules*, 2nd Edn (London: Chapman and Hall).
Gaydon, A. G., and Penney, W. G., 1945, *Proc. Roy. Soc.* A, **183,** 374.
Hagstrum, H. D., 1947 a, *Phys. Rev.*, **71,** 376; 1947 b, *Ibid.*, **72,** 947; 1951, *Rev. Mod. Phys.*, **23,** 185; 1953, *Mass Spectroscopy in Physics Research* (Washington: Nat. Bur. Stand.), Paper 26.
Hagstrum, H. D., and Tate, J. T., 1941, *Phys. Rev.*, **59,** 354.
Hall, G. G., 1951, *Proc. Roy. Soc.* A, **205,** 541; 1953, *Trans. Faraday Soc.*, **49,** 113; 1954, *Ibid.*, **50,** 319.
Hall, G. G., and Lennard-Jones, Sir John, 1950, *Proc. Roy. Soc. A*, **202,** 155.
Harrison, M. A., and Geballe, R., 1953, *Phys. Rev.*, **91,** 1.
Herzberg, G., 1937, *Chem. Rev.*, **20,** 145; 1939, *Molecular Spectra and Molecular Structure* (New York: Prentice-Hall); 1950, *Spectra of Diatomic Molecules* (New York: Van Nostrand Co. Inc.)
Hickam, W. H., Fox, R. E. and Kjeldaas, T., 1954, *Phys. Rev.*, **96,** 63.
Hipple, J. A., and Condon, E. U., 1945, *Phys. Rev.*, **68,** 54.
Hipple, J. A., Fox, R. E., and Condon, E. U., 1946, *Phys. Rev.*, **69,** 347.
Hipple, J. A., and Stevenson, D. P., 1943, *Phys. Rev.*, **63,** 121.
Honig, R. E., 1948, *J. Chem. Phys.*, **16,** 105.
Institute of Petroleum, 1952, *Mass Spectrometry;* 1954, *Applied Mass Spectrometry.*
Jevons, W., 1936, *Proc. Roy. Soc.* A, **48,** 563.
Jones, T. J., 1927, *Phys. Rev.*, **55,** 203.
Kallmann, H., and Rosen, B., 1930, *Z. Phys.*, **61,** 61.
Kirkbride, F. W., and Davidson, F. G., 1954, *Nature, Lond.*, **79,** 174.
Kistiakowsky, G. B., Knight, H. T., and Malin, M. E., 1951, *J. Amer. Chem. Soc.*, **73,** 2972.
Kistiakowsky, G. B., and Van Artsdalen, E. R., 1944, *J. Chem. Phys.*, **12,** 469.
Kusch, P., Hustrulid, A., and Tate, J. T., 1937, *Phys. Rev.*, **52,** 840, 843.
Laidler, K. J., 1954, *J. Chem. Phys.*, **22,** 1740.
Langer, A., and Hipple, J. A., 1946, *Phys. Rev.*, **69,** 691.
Lawrence, E. O., 1926, *Phys. Rev.* **28,** 947.
LeGoff, P. J., 1953, *J. Chim. Phys.*, **50,** 423.
Lennard-Jones, Sir J. E., 1949, *Proc. Roy. Soc.* A, **198.**
Long, L. A., 1949, *Proc. Roy. Soc.* A, **198,** 62.
Long, L. A., and Norrish, R. G. W., 1946, *Proc. Roy. Soc.* A, **187,** 337.
Lossing, F. P., Ingold, K. U., and Henderson, I. H. S., 1954 a, *Applied Mass Spectrometry* (London: Institute of Petroleum) p. 102; 1954 b, *J. Chem. Phys.*, **22,** 621.
Lossing, F. P., and Tickner, A. W., 1952, *J. Chem. Phys.*, **20,** 907.
Lossing, F. P., Tickner, A. W., and Bryce, W. A., 1951, *J. Chem. Phys.*, **19,** 1254.
Lozier, W. W., 1930, *Phys. Rev.*, **52,** 843; 1933, *Ibid.*, **44,** 575; 1934, *Ibid.*, **46,** 268.
Mann, M. M., Hustrulid, A., and Tate, J. T., 1940, *Phys. Rev.*, **58,** 340.
Marriott, J., 1954, *Thesis*, University of Liverpool.
Marriott, J., and Craggs, J. D., 1953, *E.R.A. Report No. L/T/301*; 1954 a, *Ibid.* L/T/308; 1954 b, *Applied Mass Spectrometry* (London: Institute of Petroleum), p. 173.
Marriott, J., Thorburn, R., and Craggs, J. D., 1954, *Proc. Phys. Soc.* B, **67,** 437.
Massey, H. S. W., 1950, *Negative Ions* (Cambridge: University Press).
Massey, H. S. W., and Burhop, E. H. S., 1952, *Electronic and Ionic Impact Phenomena* (Oxford: University Press).
McDowell, C. A., 1954 a, *Applied Mass Spectrometry* (London: Institute of Petroleum), p. 129; 1954 b, *Trans. Faraday Soc.*, **50,** 423.
McDowell, C. A., and Cox, B. C., 1952, *J. Chem. Phys.*, **20,** 1496; 1954, *Ibid.*, **22,** 946.
McDowell, C. A., and Warren, J. W., 1951, *Disc. Faraday Soc.*, **10,** 53; 1952, *Trans. Faraday Soc.*, **48,** 1084.
Meek, J. M., and Craggs, J. D., 1953, *Electrical Breakdown of Gases* (Oxford: University Press).

MORRISON, J. D., 1953, *J. Chem. Phys.*, **21**, 1767; 1954, *Ibid.*, **22**, 1219.
MORRISON, J. D., and NICHOLSON, A. J. C., 1952, *J. Chem. Phys.*, **20**, 1021.
MULLIKEN, R. S., 1933, *Phys. Rev.*, **43**, 279; 1935 a, *J. Chem. Phys.*, **3**, 517; 1935 b, *Phys. Rev.*, **47**, 413; 1949 *J. Chim. Phys.*, **46**, 497, 675.
MULLIKEN, R. S., RIEKE, C. A., and BROWN, W. G., 1941, *J. Amer. Chem. Soc.*, **63**, 41.
NATIONAL BUREAU OF STANDARDS, 1953, *Mass Spectroscopy in Physics Research* (Washington: National Bureau of Standards).
NEUERT, H., 1952, *Z. Naturf.*, **7a**, 293; 1953, *Ibid.*, **8a**, 459.
NOTTINGHAM, W. B., 1939, *Phys. Rev.*, **55**, 203.
OSBERGHAUS, O., 1950, *Z. Phys.*, **128**, 366.
PARR, R. G., and TAYLOR, G. R., 1951, *J. Chem. Phys.*, **19**, 497.
PARTI, Y. P., and SAMUEL, R., 1937, *Proc. Phys. Soc.*, **49**, 568.
PAULING, L. C., 1940, *The Nature of the Chemical Bond* (New York: Cornell University Press).
POPLE, J. A., 1953, *Trans. Faraday Soc.*, **49**, 1375.
PRICE, W. C., 1936, *J. Chem. Phys.*, **4**, 539.
ROBERTSON, A. J. B., 1952, *Mass Spectrometry* (London: Institute of Petroleum), p. 47; 1954, *Applied Mass Spectrometry* (London: Institute of Petroleum), p. 112.
SCHAEFFER, O. A., and HASTINGS, J. M., 1950, *J. Chem. Phys.*, **18**, 1048.
SHULER, K., 1953, *J. Chem. Phys.*, **21**, 624.
SMITH, L., and BLEAKNEY, W., 1936, *Phys. Rev.*, **49**, 883.
SMYTH, H. D., and STUECKELBERG, E. C. G., 1930, *Phys. Rev.*, **36**, 478.
SPONER, H., 1935, *Molekülspektren* (Berlin: Springer).
STEVENSON, D. P., 1942, *J. Chem. Phys.*, **10**, 291; 1947, *Ibid.*, **15**, 409; 1950, *Ibid.*, **18**, 1347; 1951, *Disc. Faraday Soc.*, **10**, 35; 1953, *Trans. Faraday Soc.*, **49**, 867.
SUGDEN, T. M., and PRICE, W. C., 1948, *Trans. Faraday Soc.*, **44**, 108.
SZWARC, M., 1950, *Chem. Rev.*, **47**, 75.
TATE, J. T., and SMITH, P. T., 1932, *Phys. Rev.*, **39**, 270.
TATE, J. T., SMITH, P. T., and VAUGHAN, A. L., 1935, *Phys. Rev.*, **48**, 525.
THORBURN, R., 1954, *Applied Mass Spectrometry* (London: Institute of Petroleum), p. 185.
TICKNER, A. W., BRYCE, W. A., and LOSSING, F. P., 1951, *J. Amer. Chem. Soc.*, **73**, 5001.
VOUGHT, R. H., 1947, *Phys. Rev.*, **71**, 93.
WALDRON, J. D., 1954, *Trans. Faraday Soc.*, **50**, 102.
WALDRON, J. D., and WOOD, K., 1952, *Mass Spectrometry* (London: Institute of Petroleum), p. 16.
WANNIER, G. H., 1953, *Phys. Rev.*, **90**, 817.
WARREN, J. W., and CRAGGS, J. D., 1952, *Mass Spectrometry* (London: Institute of Petroleum), p. 36.

A SURVEY OF FIELD THEORY

By R. E. PEIERLS, A. SALAM,† P. T. MATTHEWS and G. FELDMAN
University of Birmingham
† St. John's College, Cambridge

Lectures given at the Autumn Meeting of the Physical Society in the University of Birmingham
13th and 14th December 1954

CONTENTS

PART I. GENERAL BACKGROUND
By R. E. PEIERLS

PAGE
§ 1. Introduction 424
§ 2. The beginnings of field theory 424
§ 3. Interaction between field and matter. Difficulties 427
§ 4. Resolution of the difficulties. Renormalization 430

PART II. FIELDS AND PARTICLES
By A. SALAM

PAGE
§ 1. Introduction 433
§ 2. The classical radiation field 433
§ 3. Quantum theory 433
§ 4. Quantization of the matter field 436
§ 5. The Dirac equation 438
§ 6. General field theory 440
Appendix 440

PART III. INTERACTIONS OF QUANTIZED FIELDS
By P. T. MATTHEWS

PAGE
§ 1. Annihilation and creation operators 441
§ 2. Electrodynamics 442
§ 3. Scattering phenomena 444
§ 4. Anomalous electron moment and the Lamb shift 446
§ 5. Renormalization 446
§ 6. Meson theories 448
§ 7. Possible interactions 450

PART IV. PROPERTIES OF THE π-MESON
By P. T. MATTHEWS

PAGE
§ 1. π and μ Mesons 452
§ 2. Spin of π^0 453
§ 3. Spin of $\pi^\pm$ 454
§ 4. Parity of $\pi^\pm$. The Panofsky experiment 455
§ 5. Charge independence 456
§ 6. Isotopic spin 457
§ 7. The meson–nucleon interaction 460

PART V. HIGH ENERGY EXPERIMENTS 462
By G. FELDMAN

PART VI. FINAL REMARKS. THE NEW PARTICLES
By R. E. PEIERLS

PAGE
§ 1. Present state of electromagnetic theory 470
§ 2. Note on the classical limit for waves and particles 471
§ 3. Difficulties of meson theory 473
§ 4. New particles of surprisingly long life 474
§ 5. Different types of coupling 475

Abstract. These six articles attempt to summarize the concepts and results of the field theory of elementary particles in terms which, it is hoped, will be understood by experimentalists or theoretical physicists working in other fields.

Part I is a general, partly historical, introduction. The concept of virtual states and the physical ideas underlying the renormalization of mass and charge are explained.

The outline of the theory starts in Part II, which deals with the description of a quantized field—the quantized electromagnetic field for example—and its relation to an assembly of particles.

Part III describes the interactions of such particles. It is shown how solutions to problems in electrodynamics (the interaction of electrons and positrons with photons) may be obtained as a power series in the fine structure constant $e^2/4\pi\hbar c(=1/137)$.

A number of possible meson–nucleon interactions are then developed by analogy with electrodynamics. To decide which of these actually occurs in nature, it is necessary to determine from experiment the spin and intrinsic parity of the π-meson. Part IV is devoted to this problem. The charge independence of nuclear forces expressed in terms of isotopic spin is then used to determine, as uniquely as possible, the π-meson–nucleon interaction in qualitative agreement with experimental data. Quantitative agreement has not yet been obtained owing mainly to the largeness of the coupling constant.

In Part V the experimental evidence on π-mesons is reviewed. It is emphasized that a variety of effects can be explained on the assumption that there is a short-lived excited state of the nucleon.

Part VI starts with a discussion of the classical limit of the electromagnetic and Dirac fields and concludes with some considerations of the properties of the recently discovered heavy mesons and some general regularities of the couplings between them.

PART I. GENERAL BACKGROUND

By R. E. Peierls

§ 1. Introduction

The purpose of this and the following articles is to provide a brief account of the ideas of the quantum theory of fields for the non-specialist†. In such a brief survey one cannot escape over-simplification, and many refinements which have been of vital importance to the progress of the subject over the last few years had to be omitted. It has been our intention, however, to simplify without misleading, and we have attempted to indicate clearly the points where there remain difficulties and uncertainties, and to give a balanced view of the extent to which a clear understanding of the problems exists at the present time.

§ 2. The Beginnings of Field Theory

The application of quantum theory to a radiation field and in particular to electromagnetic radiation is as old as quantum theory itself. Indeed the first introduction by Planck of the constant h related to light quanta, i.e. to the dynamics of radiation, and it was only later that Niels Bohr extended these ideas to the dynamics of particles.

The further development of the subject then concentrated on the description of particles and systems of particles, and this led to the development of the present form of quantum mechanics following the ideas of de Broglie, Heisenberg, Schrödinger and Dirac.

† A more detailed mathematical treatment of the same topics can be found in the article by J. C. Gunn in this volume (p. 127).

Since most of the empirical evidence which guided this development came from spectroscopy, problems involving the emission and absorption of light by atomic systems always played a vital part in this work. Their discussion, however, did not require a formal quantum description of the electromagnetic field, for the following reason: If one is concerned with the absorption of light, say by an atom, one may consider not the realistic case of a weak radiation field, for which the photon concept is essential, but instead the hypothetical case of an extremely intense field in which the radiant energy is so high (i.e. in which so many photons are present) that the discrete quantum nature is of no consequence. In this case one may treat the electromagnetic field as if it were a given classical external field, and the vector and scalar potential of this field then appear in the Schrödinger equation of the atom in the same manner in which the Coulomb potential of the atomic nucleus may be introduced as a given function of space and time. One can thus obtain the probability of excitation, in which the atom receives energy from the external field, as a function of the intensity of the light wave, and one will at the same time find that there exists a probability of the inverse transition in which the atom loses energy. This latter process, which has to be interpreted as ' induced emission ' is also proportional to the light intensity. It is then possible to obtain the physically interesting spontaneous emission by using the Einstein relations between absorption induced and spontaneous emission. According to these relations we may express the probability of absorption in terms of the number N of photons in the light wave (the energy of one particular mode of the radiation field divided by the energy $h\nu$ of one photon) and it will turn out to be proportional to N. The probability of emission is then correctly obtained if we replace N by $N+1$, and in this result the term with N represents the induced and the 1 the spontaneous emission.

The justification of these relations comes from the knowledge of the statistical equilibrium of the radiation field at a given temperature, expressed in Planck's law, and this in turn relies on the application of quantum ideas to the radiation field. It may therefore be said that from the very beginning the quantum treatment of the field was inherent in the development of quantum mechanics. However, the method of reasoning sketched here shows that no detailed mathematical description of the radiation field had to be included in the formalism.

It was, however, also clear that there existed finer features of the interaction between radiation and matter which could not be dealt with in such a simple way, and that these would require an extension of the methods of quantum theory to the electromagnetic field. How to do this was shown first by Dirac, who solved the problem for the vacuum field, i.e. the electromagnetic field in the absence of any absorbing or emitting atomic systems. This generalization of the ideas is really quite straightforward and it can best be explained by comparison with a rather simpler problem, i.e. the vibrations of an elastic body.

Consider for the moment an elastic solid built of atoms or, for the sake of further simplicity, consider a linear chain of atoms connected by forces which tend to hold these atoms at regular intervals from each other. It is well known that, if the displacements of the atoms from their equilibrium positions are small, this leads to coupled linear equations of motion for the atoms, and that the kinetic and potential energy of the atoms are quadratic functions of their velocities and positions respectively. It is also known that in classical mechanics the solution of this problem can be obtained in terms of normal coordinates and normal modes of oscillation.

If the chain is long enough for end effects to be unimportant, or if we consider an endless chain by placing our atoms on a circle rather than in a straight line, each normal mode of motion is of the form

$$u_n = a \cos knd;$$

here u_n is the displacement of the nth atom from its equilibrium position, k is the wave vector (2π times the inverse wavelength) characteristic of the normal mode in question, d the distance between the atoms. a is the amplitude of the appropriate mode. It is often more convenient to introduce progressive waves rather than standing waves as specified here, but for the present purpose the distinction is unimportant.

The most general configuration of atoms can be expressed as a superposition of such normal modes, and thereby the amplitudes a_k of all the modes may replace the atomic coordinates as variables. Since the relation between the coordinates and the amplitudes a_k is linear, this transformation retains its validity in quantum mechanics and we can take over the classical equations of motion for the amplitudes a_k which are now independent equations, each relating only to one amplitude. Each of these equations is then the equation of simple harmonic motion, with a frequency whose relation to the wave vector is known if we know the force constants. For long waves in particular the ratio of the frequency to the wave vector is the velocity of sound.

The quantum laws then immediately show that we have to introduce a quantum number $N(k)$ for each of the normal modes, and the state of the whole system is completely specified by specifying all these quantum numbers. The state of lowest energy is clearly obtained by choosing all these quantum numbers to be zero (all oscillators in their ground state). We may next consider a group of states in which only one of the normal modes has $N=1$, the rest still zero. Instead of saying that one of the oscillators is excited by one quantum we may specify this state by saying we are dealing with one quantum of oscillation, or one ' phonon ', and then further specify the wave vector k of this phonon for a complete description of the state.

Similarly any other state of vibration may be described by first specifying the total number of phonons, i.e. the sum of all the quantum numbers of the oscillators, and further the wave vectors of all these phonons, of which some may be equal, if some of the oscillators are excited to states of quantum numbers greater than 1.

If we are concerned only with waves of such great wavelength that the atomic structure of the chain is unimportant, we have here essentially the theory of an elastic continuum. The purpose of the simple picture discussed here is to point out that, starting from a description in terms of the atomic displacements, which in the limit of a continuum becomes the theory of elastic waves, we are naturally led without the use of any new concepts to the idea of phonons, i.e. the idea of discrete particles, and this is a particularly transparent example of the relation of the particle and the wave aspect.

The situation does not change at all if one considers the limit of infinitely small spacing between the atoms, so that the discrete equations of motion are replaced by the differential equations of an elastic continuum, and the same ideas can be applied to the equations of the electromagnetic field without essential change. The only new feature of the continuous case is that there is now no natural upper limit to the wave vector, or lower limit to the wavelength. In the normal modes describ-

ing atomic motion there is no sense in defining waves of wavelengths less than the atomic distance, but in the continuum we must define an infinite sequence of normal modes, corresponding to the fact that a continuous field has an infinite number of degrees of freedom. As long as we are concerned only with the electromagnetic field in the absence of charged particles this change causes no serious difficulty. It is true that quantum mechanics predicts a 'zero-point' energy for each oscillator, even in the lowest state, and if we are dealing with an infinite number of oscillators this zero-point energy would have to be infinite. This is equivalent to saying that the electromagnetic field at any point in space has infinite fluctuations even in the lowest state, i.e. in the absence of any photons. However, no physical experiment is ever concerned with the field at a mathematical point. If one discusses measurements of the force on some physical test object, which depends on the average value of the field over finite region in space and time, then the fluctuations are finite and, as has been shown by an analysis due to Bohr and Rosenfeld, correspond to just what is physically required from the fact that in the course of the measurement the test body must be capable of emitting photons whose energy and momentum must, by the uncertainty principle, be unknown if one is to obtain a measurement of the field.

Another consequence of the zero-point energy is that the calculation of the energy content of the radiation field would give an infinite answer even in the case of no photons. However, this apparent difficulty is easily avoided since it is always possible to readjust the zero from which energy is measured, and we are concerned physically only with the energy differences between states containing different numbers of photons or photons in different modes of motion. We shall see, however, that the presence of an infinite number of degrees of freedom does lead to difficulties in the treatment of interaction between field and charged particles.

§ 3. Interaction between Field and Matter. Difficulties

The next step is to extend the quantum description so as to include the interaction of charged particles with the field. This must, in the first place, account for the emission and absorption of radiation by atoms and other systems. As has been mentioned previously the emission and absorption probabilities which must come out from such a description are known in advance from simple and general principles. The equations required for this purpose were first set up by Dirac (1927) and further developed by Heisenberg and Pauli (1929), Fermi (1932) and others; Weisskopf and Wigner (1930) used the same principle to solve the problems of emission and absorption, and they also discussed radiation damping, i.e. the natural width of spectral lines and other subtler effects.

At this point a new difficulty is met for the first time.

In order to understand the nature of this difficulty we must first of all understand that the interaction between an atom and the electromagnetic field is extremely weak. For example the probability of emission of a photon from an excited state of hydrogen is of the order of

$$\frac{e^2}{4\pi\hbar c}\left(\frac{v}{c}\right)^2$$

during one (classical) period of revolution. Here $e^2/4\pi\hbar c$ is the famous fine-structure constant ($\hbar$ stands for $h/2\pi$)*, which is numerically 1/137, v the electron

* Readers will be more familiar with this in the form $e^2/\hbar c$; unfortunately the use of Heaviside units has become a firm convention in the literature on this subject.

velocity and c the velocity of light. In a low excited level of hydrogen v/c is again of the order of the fine-structure constant, so that on the average the electron performs 137^3 revolutions before on the average it makes a spontaneous transition with the emission of light. In other words the natural width of the hydrogen lines is of the order of $1/137^3$ of the line frequency.

This situation makes it reasonable to regard the coupling between the atom and the radiation field as a very small effect and treat it only to a first-order approximation. This amounts to saying that starting from the states of the atom in the absence of the coupling it is a good approximation to consider only the possibility of the emission of radiation and to neglect the reaction of the emitted radiation on the atom. In so doing we neglect the radiative broadening of the line which is due to this reaction and as we have seen this is indeed an extremely small effect.

But we must take into account the second-order effects if we wish to calculate radiative broadening. We also must do so if we wish to calculate the force between different charged particles from the equations of the electromagnetic field to which this also of course owes its existence. By far the greatest part of the force is usually the static Coulomb interaction, and this is described by a ' longitudinal ' electric field, i.e. a field of which curl $\mathbf{E}=0$ everywhere, but div $\mathbf{E}$ does not vanish everywhere. Such longitudinal fields are not affected by the quantum nature of the field, but it is still desirable to include them in the formal description of the field, both for the sake of uniformity and in particular to bring out clearly the relativistic features, because the distinction between the longitudinal electric and the transverse magnetic field is not independent of the state of motion of an observer. In any event it is necessary to use the quantum field concept if one wants to include also the magnetic interaction between particles which become appreciable at high velocities.

These effects also depend on the reaction on an electron of the field produced by this same or another electron and therefore require a treatment of the coupling at least to second order.

It is a general feature of quantum mechanics that the effect of small coupling terms in the second and higher order is described in terms of successive ' virtual ' transitions through intermediate states of the combined system. For example both the radiative damping and the electromagnetic interaction between charged particles is described as the ' virtual ' emission of a photon by one of the particles and the subsequent reabsorption of the photon by the same or another particle. The term ' virtual ' here serves as a reminder that we have to consider intermediate states in which energy is not conserved. The apparent contradiction can be understood if we remember that by the uncertainty principle the energy of a state is not observable if the state exists only for a certain time. More precisely, if a state exists only for a time Δt then its energy cannot be known to better than ΔE, where $\Delta E \Delta t$ must be at least of the order of the quantum constant $\hbar$. Crudely speaking, a system may borrow an amount of energy for a short time, provided the loan is returned before it is possible to discover that the amount is missing. For example, the well-known phenomenon of the transition of particles through potential barriers, which is important in the Gamow theory of alpha-decay, can be described simply in these terms. The particle can get through the potential barrier notwithstanding the lack of required kinetic energy, provided the energy is balanced again soon enough to prevent an actual verification of the temporary lack of balance. Qualitatively this explains that the transition becomes more difficult if the barrier

is higher (therefore requiring a bigger loan) or if it extends further (so that it becomes more difficult to restore the balance in a short time).

Considering the interaction of two electrons we have then to envisage first the emission of a photon by one of them; this cannot take place as a real process since it is well known that the emission of one photon by a free particle is never compatible with conservation of energy and momentum. This is followed by the reabsorption of the photon by the other particle, thereby restoring the balance. The photon in question may have any energy, but if its energy is high it can be effective only if it is reabsorbed after a short time, and since actions are not propagated with a speed greater than that of light, this in turn is possible only if the second particle is not too far away. In other words a photon of energy E can play a part only if the distance between two electrons is less than h/Ec. The smaller the distance r between the particles, the greater is therefore the range of photon energies which have to be allowed for; this argument can be made quantitative so as to show the interaction energy as proportional to the inverse distance.

But by the same mechanism even a single electron will acquire an additional energy due to its interaction with the radiation field. The virtual process to be considered here is the emission and reabsorption of a photon by the same electron. In this case no distance is involved (as long as the electron is treated as a point charge) and consequently virtual photons of all possible energies contribute. As a result the change in the energy, the so-called self-energy of the electron, becomes infinite.

This result is not unexpected since it corresponds to a fact which is well known in the classical theory of the electron. If we compute the total energy content of the classical electrostatic field of a point charge, it is also infinite, as may be seen from the fact that the energy of the field surrounding a uniformly charged sphere of total charge e and radius r is $\frac{3}{5}e^2/r$ and therefore tends to infinity as r tends to 0.

It was, in fact, believed that a substantial part of the whole of the rest energy of the electron was due to this cause, and in order to make this self-energy equal to mc^2 one would have to assume the electron to be a sphere of radius r about e^2/mc^2.

We also know that in quantum theory the electron has a spin and, associated with it, a magnetic dipole moment. The field of a dipole is more strongly divergent and if this magnetic self-energy is included, the ' radius ' of the electron would have to be assumed even larger. This result is also reproduced in quantum theory if one uses the relativistic equation of Dirac for the electron which does contain the spin and the corresponding magnetic moment.

However, the description of the single electron by Dirac's equation contains predictions which are quite unphysical, since it contains the possibility of the electron moving in states of negative kinetic energy. It is well known that this difficulty is overcome only by assuming that all states of negative energy are normally filled, a vacancy in such a negative energy state manifesting itself as a positron. Now in the calculation of the self-energy which I have sketched (both the electrostatic and the magnetic part) a large part is played by virtual processes in which the electron makes a transition to a state of negative energy. If these states are filled, such transitions cannot take place; on the contrary the presence of the electron inhibits also virtual processes which could take place in the vacuum in its absence. These are transitions in which an electron is raised from a negative to a positive energy state, in other words the virtual creation of electron pairs. Such terms occur in the self-energy of the vacuum and, since the observable energy of an

electron is the difference between the state with the electron present and with it absent, the suppression by the electron of part of the self-energy of the vacuum means a negative contribution to its self-energy. If this is worked out quantitatively, the total self-energy of the electron is much less divergent than in the one-particle description and diverges only logarithmically. To limit the self-energy to a reasonable value the electron radius may therefore be taken as small as something like $(\hbar/mc)e^{-137}$. There is no reason whatever to believe in the validity of our present concepts for distances so fantastically small.

However, the attempt to associate a finite size with the electron has so far failed because of the difficulty of reconciling a rigid structure with the principle of relativistic invariance. As long as we could not avoid the treatment of the electron as a point charge all second-order calculations tended to contain infinite terms which, even though they were now only logarithmically infinite, made the theory quite useless for practical purposes.

One's first idea was to see if this infinite term was just an additive constant, such as was the case with the infinite zero-point energy of the vacuum field which was mentioned earlier. It was found, however, that this was not so and that even the differences between the energies of different states of motion of the electron if calculated in this manner turned out to be infinite.

§ 4. Resolution of the Difficulties. Renormalization

As a result of this difficulty the position of the theory just after the war was that each physical quantity could be calculated to the first order in which the effect in question occurred, but any attempt to calculate corrections of higher order would give an infinite answer which it was impossible to interpret. The incentive to the further development was provided by the discovery due to Lamb and Retherford of anomalies in the levels of the hydrogen atom, and in the process of trying to account for these discrepancies a method was developed which helped to overcome the difficulty caused by the infinities. The physical idea behind this was first suggested by Kramers, and the mathematical formalism to make it precise was developed by Tomonaga (1946) and Schwinger (1948).

The basis of this idea lies in the remark that the nature of the self-energy is to give the electron a greater effective mass. If the electron is accelerated the field which it carries is accelerated with it and also contributes to its kinetic energy, by the general principle of relativity. It is therefore natural that the self-energy should also affect the energy differences between different states of motion of the electron. Now whenever we measure the mass of the electron we include the self-mass caused by its electromagnetic field. We must therefore distinguish between the mass of a ' bare ' electron, which it would have if stripped of its interaction with the field, and its real observable mass m. We have in fact $m=m_0+\delta m$ if δm is the electromagnetic self-mass. If in the mathematical treatment we include the electromagnetic effects then we must start from an electron of mass m_0. There is no empirical way of finding m_0, but we must determine it in such a way that after including the electromagnetic effects the total mass m is given correctly. The process of adjusting the constants to allow for this difference is known as ' renormalization '.

It is now found that even though the electromagnetic mass comes out infinitely large in the equations and therefore m_0 also has to be infinite, calculations of any

observable effect give a finite answer provided one ensures that the real mass m has the correct (and finite) value.

In these terms it is possible to understand the origin of the anomaly in the hydrogen spectrum, the so-called Lamb shift, in this way. Both if the electron is free and if it is moving in an orbit around the proton it continuously emits and reabsorbs virtual photons and as we have seen in both cases this results in an infinite correction to its mass. We have to arrange, however, that the resultant mass is finite. It will be independent of the velocity of the electron provided it is moving uniformly, since the principle of relativity, with which our theory is compatible, guarantees that a state of uniform motion cannot make any change.

A difference arises, however, when the electron is in accelerated motion, and we may picture this by saying that the electron has emitted a virtual photon and then changes its velocity slightly under the influence of the external field before it has reabsorbed it. There results therefore a change in the effective mass of the electron which depends on the force acting on it. This change is finite because it comes predominantly from virtual photons of moderate energy. If the photon has very high energy, then by the uncertainty principle it must be emitted and reabsorbed after a very short interval of time, and during such a short interval of time the electron has no opportunity of having changed its velocity much under the effect of the field force. It follows that the correction will be less divergent than the self-energy itself and it is not surprising that it should be finite.

Since then the calculation of the Lamb shift has been made extremely precise and the measurement has also reached the same impressive accuracy (see Lamb 1951). The agreement is almost perfect.

A general proof that calculations using the renormalization procedure will always give finite answers for any observable effect was given by Dyson (1949), using the elegant mathematical method developed by Feynman (1949).

In dealing with such problems one also has to take into account that in the theory including electron–positron pairs the vacuum behaves like a dielectric in so far as the presence of any electromagnetic field gives rise to virtual creation and annihilation of electron pairs. This ' vacuum polarization ' means that any charge in the field is in part compensated by charges induced in the vacuum in the form of virtual pairs. And it turns out again that this vacuum polarization effect is divergent. The remedy is the same as before. We must remember that we never see a charge without the accompanying vacuum effects and that it is the net charge after allowing for these effects which we observe. Whereas the total induced charge cannot by any observation be separated from the intrinsic charge of a particle, its distribution in space can be observed. Part of it is not localized on the charge itself, but spread over a region of the order of the Compton wavelength $\hbar/mc$ around it. This part is finite and small, but it has to be allowed for both in the calculation of the Lamb shift and in the energy levels of the ' mesic atoms ', i.e. the short-lived systems formed by a negative meson revolving in an orbit around an atomic nucleus, and the effect is important for recent experiments on such systems.

Another interesting phenomenon which is now fully understood is the anomalous magnetic moment of the electron, a slight deviation from the Bohr magneton, which Dirac's equation predicts. This again can be understood in terms of the virtual emission and reabsorption of photons, if the divergent, but unobservable self-energy effects are eliminated by renormalization. It will be discussed in more detail in Part III.

We have thus seen that the application of the new methods of quantum field theory have been successful in showing that the apparent infinities in the theory arise, as far as physical predictions go, only from faulty interpretation, that for most purposes the good agreement between the simple first-order calculations with experiment is satisfactory because the corrections predicted by the theory are indeed small, and that moreover when the accuracy of measurement permits a determination of these corrections they are found to be in agreement with the theory. In this it has been most important that the fine-structure constant is a very small number. This has made the procedure of successive approximations reasonable, and it means, for example, that while the possibility of the emission of a large number of virtual photons is of importance for the self-mass of the electron, they will mostly be of high energy, hence of short duration and unaffected by the details of the motion. For the consideration of the Lamb shift and other such corrections we may restrict ourselves to the consideration of photons of which there is at most one at a time. The chance of there being two of them is small, of the order of one part in 137.

Similar problems arise when we consider the interaction of nucleons (i.e. electrons or protons) with the meson field. Here again an interaction is caused by the virtual emission of a meson by one particle and the reabsorption by another. It is at once evident that this will now lead to a short range force, since even the slowest meson has an energy of μc^2 (μ being the mass of a meson) and can therefore not persist for a time longer than about $h/\mu c^2$. Since no action can be transmitted with a velocity greater than that of light such mesons can therefore not cause an appreciable force over distances greater than $h/\mu c$. This is in simplest terms the basis of Yukawa's postulate of mesons of a few hundred times the electron mass to account for the range of nuclear forces. It will be seen later, however, that the quantity corresponding to the fine structure constant in the interaction between nucleons and mesons is not small. While this makes the calculation of the more complicated interactions of much greater importance it also makes it much more difficult. The following parts will explain the language in which one has to discuss these problems and indicate the progress which has been made.

Part II. FIELDS AND PARTICLES

By A. Salam

§ 1. Introduction

Quantum theory of fields has developed from two distinct points of view. On the one hand the basic concepts of the theory arise naturally when an attempt is made to combine the wave and particle aspects of radiation. On the other hand, they arise equally naturally when we try to write a relativistic Schrödinger equation for a free particle. We wish to trace the development of field theory from these two points of view.

§ 2. The Classical Radiation Field

We begin by considering the classical electromagnetic field. Faraday and Maxwell pictured all space surrounding electric charges as filled with electric and magnetic fields, characterized by vectors $\mathbf{E}$ and $\mathbf{H}$. The fields are not introduced as mere mathematical accessories; they are dynamical systems in their own right: so much so, that the theory specifies, at each point of space and time, the energy and the momentum density of the fields. The total energy is

$$H=\frac{1}{2}\int(\mathbf{E}^2+\mathbf{H}^2)d^3x, \qquad \ldots\ldots(2.1)$$

and the total momentum

$$\mathbf{G}=\frac{1}{c}\int(\mathbf{E}\times\mathbf{H})d^3x. \qquad \ldots\ldots(2.2)$$

The vectors $\mathbf{E}$ and $\mathbf{H}$ satisfy Maxwell's equations. If scalar and vector potentials V and $\mathbf{A}$ are introduced, such that $\mathbf{E}=-\text{grad}\ V-(1/c)d\mathbf{A}/dt$ and $\mathbf{H}=\text{curl}\ \mathbf{A}$, Maxwell's equations at large distances from the charged particles producing the fields reduce to

$$\nabla^2\mathbf{A}-\frac{1}{c^2}\frac{d^2\mathbf{A}}{dt^2}=0, \qquad \ldots\ldots(2.3)$$

$$\nabla^2V-\frac{1}{c^2}\frac{d^2V}{dt^2}=0, \qquad \ldots\ldots(2.4)$$

and

$$\text{div}\ \mathbf{A}+\frac{dV}{cdt}=0. \qquad \ldots\ldots(2.5)$$

The solutions of these equations describe the familiar propagation of electromagnetic waves. For a pure radiation field we can in fact introduce a further simplification and set $V=0$.

§ 3. Quantum Theory

So much for the classical theory. Now, one of the basic hypotheses of quantum theory is the quantum hypothesis of Planck and Einstein: that energy associated

with a plane electromagnetic wave of frequency $\omega(=2\pi\nu)$ exists in units of $\hbar\omega$ ($\hbar=h/2\pi$). Each unit, called a photon, acts as a discrete particle of energy $\hbar\omega$ and momentum $\hbar\boldsymbol{\omega}/c$. The basic problem of quantum theory is to bring out this quantum character of radiation—to try to express the classical radiation field as an assembly of photons. Mathematically, Planck's hypothesis amounts to the requirement that H and $\mathbf{G}$ can be written as

$$H=\sum_{\omega} n_{\omega}\hbar\omega, \qquad \ldots\ldots(2.6)$$

$$\mathbf{G}=\sum_{\omega} n_{\omega}\frac{\hbar\boldsymbol{\omega}}{c}, \qquad \ldots\ldots(2.7)$$

where the integer n_{ω} is the number of photons of frequency ω. Dirac (1927) made the first attempt to solve this problem, making use of precisely the same methods as he had developed for atomic and molecular systems. The theory starts with just two assumptions:

(i) The field quantities, $\mathbf{E}$, $\mathbf{H}$, $\mathbf{A}$, V or H and $\mathbf{G}$ are not classical quantities, but linear operators.† The eigenvalues of the energy and the momentum operators give the possible results of actual measurements of these quantities.

(ii) We postulate that the operators $\mathbf{E}$ and $\mathbf{H}$ do not commute; that $\mathbf{E}_x\,\mathbf{H}_x$, for example, is not necessarily equal to $\mathbf{H}_x\,\mathbf{E}_x$.

In non-relativistic quantum mechanics, if x is the position operator for a particle and p the conjugate operator, specifying momentum, it is well known that $xp\neq px$. Specifically,‡

$$xp-px=i\hbar. \qquad \ldots\ldots(2.8)$$

In exact analogy with this, we postulate, that, for example, $A_x(\mathbf{r},\,t)$ and its time derivative $(d/dt)A_x(\mathbf{r},\,t)$ (analogue of p) do not commute, and that

$$A_x(\mathbf{r},\,t)\frac{dA_x}{dt}(\mathbf{r}',\,t)-\frac{dA_x}{dt}(\mathbf{r}',\,t)A_x(\mathbf{r},\,t)=i\hbar c^2\delta(\mathbf{r}-\mathbf{r}'). \qquad \ldots\ldots(2.9)$$

A relation of the type (2.8) implies an uncertainty in the simultaneous measurement of position and momentum, $\Delta x\,\Delta p\geqslant\hbar$. In the same way, relation (2.9) implies that a simultaneous measurement of $\mathbf{E}$ and $\mathbf{H}$ with arbitrary accuracy is not possible.

Let us now make a Fourier analysis of $\mathbf{A}$.

$$A_i(\mathbf{r},\,t)=\sqrt{\left(\frac{\hbar c}{16\pi^3}\right)}\sum_{\omega}\frac{1}{\sqrt{\omega}}[a_{\omega i}\,\exp\,(i\boldsymbol{\omega}.\mathbf{r}/c-i\omega t)+a_{\omega i}^{*}\,\exp\,(-i\boldsymbol{\omega}.\mathbf{r}/c+i\omega t)]. \qquad \ldots\ldots(2.10)$$

Since div $\mathbf{A}=0$, only two of the components of $\mathbf{A}$ are independent, corresponding to the two photon polarizations. It is easy to verify, that if $\omega=|\boldsymbol{\omega}|$, expression (2.10) for $\mathbf{A}$ satisfies Maxwell's equations. Since A_i is an operator, and no longer

† Matrices and differential operators are examples of linear operators. The meaning of the term eigenvalue will become clearer later on.

‡ If we set $p=-i\hbar d/dx$, then $xp-px=-i\hbar[x\,d/dx-(d/dx)\,x]$ operating on any function $f(x)$ just equals $i\hbar$ times $f(x)$; $-i\hbar d/dx$ is called a ' representation ' of the operator p.

a classical quantity, it is clear that the Fourier amplitudes a and a^* also are not classical numbers. In fact by a direct substitution in (2.9) we obtain

$$a_\omega a_\omega^* - a_\omega^* a_\omega = 1, \qquad \ldots\ldots(2.11)$$

which shows that a, a^* are operators which do not commute. Further,

$$H = \sum_\omega (a_\omega^* a_\omega + \tfrac{1}{2})\hbar\omega, \qquad \ldots\ldots(2.12)$$

$$\mathbf{G} = \sum_\omega (a_\omega^* a_\omega)\hbar\boldsymbol{\omega}/c. \qquad \ldots\ldots(2.13)$$

All the essential features of the quantum theory of free fields are contained in equations (2.11) and (2.12).

To relate our symbols to physical reality, all we need are the ' eigenvalues ' of $a_\omega^* a_\omega$. Once these are known, our first postulate will tell us the results of physical measurements.

A matrix representation of quantities a and a^*, satisfying equation (2.11) is easy to construct. One suitable choice is given by

$$a = \begin{pmatrix} 0 & \sqrt{1} & 0 & 0 & 0\ldots. \\ 0 & 0 & \sqrt{2} & 0 & 0\ldots. \\ 0 & 0 & 0 & \sqrt{3} & 0\ldots. \\ 0 & : & : & : & \\ : & : & : & : & \end{pmatrix}, \qquad \ldots\ldots(2.14)$$

$$a^* = \begin{pmatrix} 0 & 0 & 0 & 0 & 0\ldots. \\ \sqrt{1} & 0 & 0 & 0 & 0\ldots. \\ 0 & \sqrt{2} & 0 & 0 & 0\ldots. \\ 0 & 0 & \sqrt{3} & 0 & 0\ldots. \\ : & : & : & : & \\ : & : & : & : & \end{pmatrix}. \qquad \ldots\ldots(2.15)$$

In this representation

$$a^* a = \begin{pmatrix} 0 & 0 & 0 & 0 & 0\ldots. \\ 0 & 1 & 0 & 0 & 0\ldots. \\ 0 & 0 & 2 & 0 & 0\ldots. \\ 0 & 0 & 0 & 3 & 0\ldots. \\ : & : & : & : & \end{pmatrix}. \qquad \ldots\ldots(2.16)$$

This is a diagonal matrix. For such matrices the numbers lying along the diagonal are called ' eigenvalues '. The eigenvalues of a^*a are just the (positive) integers $n_\omega = 0, 1, 2, \ldots$

This, then, is the end of our quest. According to our first postulate a measurement of the total energy of the field must give the result $\Sigma(n_\omega + \frac{1}{2})\hbar\omega$ (the term $\Sigma\frac{1}{2}\hbar\omega$ is a constant zero-point energy of the field and can simply be ignored), while a measurement of total momentum must give the result $\Sigma n_\omega \hbar\boldsymbol{\omega}/c$. This, in fact, is Planck's hypothesis, which we have derived from Maxwell's theory, using a special (quantum) formulation for the field quantities $\mathbf{A}$, $d\mathbf{A}/dt$.

Summarizing the results so far, we started by considering the field as a dynamical system. We postulate that the field quantities are operators and that they satisfy a commutation relation, analogous to the one for position and momentum in non-

relativistic quantum theory (equation (2.9)). The field operator $\mathbf{A}$ is then Fourier analysed, the commutation relation between $\mathbf{A}$ and $d\mathbf{A}/dt$ leading to a commutation relation between the Fourier amplitudes a and a^* (2.11). All field quantities can be expressed in terms of a and a^*; in particular the total energy H and total momentum $\mathbf{G}$ of the field are expressed in terms of the operator $N_\omega = a_\omega^* a_\omega$. By explicit construction we found that the eigenvalues of N_ω are all the integers $n_\omega = 0, 1, 2, 3, \ldots$ The integers give the number of photons present for each frequency value. Once they are specified the state of the system is fully specified.

On what does the operator N_ω operate? We can construct a symbol $\Psi(\ldots, n_\omega, \ldots)$ which depends on these integers n_ω. This describes the state of the field by specifying the number of photons present of each frequency. Ψ is called the ' state function ' (or state vector). Symbolically we write

$$N_\omega \Psi(\ldots, n_\omega, \ldots) = n_\omega \Psi(\ldots, n_\omega, \ldots). \qquad \ldots\ldots(2.17)$$

So far we have found a meaning for the combination $N = a^* a$. We can in fact do more, we can assign a meaning to the quantities a and a^* themselves. From (2.11)

$$a^*(aa^* - a^* a) = a^*, \qquad \ldots\ldots(2.18)$$

that is,

$$Na^* = a^*(N+1). \qquad \ldots\ldots(2.19)$$

If Ψ is a state function describing the presence of n photons (mathematically $N\Psi = n\Psi$), then from (2.19)

$$N(a^*\Psi) = (n+1)\,(a^*\Psi). \qquad \ldots\ldots(2.20)$$

This is an equation precisely of the type (2.17). This tells us that the symbol $a^*\Psi$ describes the state of the field when $n+1$ photons are present. a^* is thus a creation operator; it creates one photon when it acts on a given state vector.

Similarly a is an annihilation operator; its operation on a state vector decreases the number of photons present by one. These creation and destruction operators are the main mathematical tools through which field theory works.

It is important to stress here that $\mathbf{A}(\mathbf{r}, t)$ is not a wave function of any particular photon. After Fourier analysis it is a sum of terms like $a_\omega \exp(i\boldsymbol{\omega}.\mathbf{r}/c - i\omega t)$. It is possible to interpret the exponential part as a wave function for an individual photon. But a and a^*, the Fourier amplitudes, are operators, and $\mathbf{A}(\mathbf{r}, t)$, as a whole, refers to the entire assembly of photons and not to any individual photon whatever. Once again, one may stress the essential simplicity of the procedure adopted. After the first two postulates are made, the rest of the picture, in terms of the creation and annihilation operators follows naturally from the logic of the mathematics.

§ 4. Quantization of the Matter Field

Let us now consider the second line of development, the quantization of what is called the matter field.

In non-relativistic theory, a particle of mass m satisfies the well-known Schrödinger equation

$$E\psi = (p^2/2m)\psi. \qquad \ldots\ldots(2.21)$$

E and $\mathbf{p}$ are operators, conventionally represented by

$$E = i\hbar\partial/\partial t, \qquad \ldots\ldots(2.22)$$

$$\mathbf{p} = -i\hbar \text{ grad}. \qquad \ldots\ldots(2.23)$$

Using these, equation (2.21) reduces to the familiar form of the wave equation:

$$i\hbar\frac{\partial}{\partial t}\psi = -\frac{\hbar^2}{2m}\nabla^2\psi. \qquad \ldots\ldots(2.24)$$

$\psi(\mathbf{r}, t)$ is called the wave function. The square of its modulus $\rho = \psi^*\psi$ gives the probability of finding the particle described by ψ, at the space point $\mathbf{r}$, at time t.

Let us now try to set up a relativistic analogue of equation (2.24). The natural generalization of (2.24) is

$$E^2\phi = (p^2c^2 + m^2c^4)\phi; \qquad \ldots\ldots(2.25)$$

on using (2.22) and (2.23) this gives

$$\nabla^2\phi - \frac{1}{c^2}\frac{d^2\phi}{dt^2} = \frac{m^2c^2}{\hbar^2}\phi. \qquad \ldots\ldots(2.26)$$

For the physical interpretation of the theory we now need some expression containing ϕ, analogous to $\psi^*\psi$, which can have the meaning of probability density. On general grounds, the only expression which may possibly serve is (see Appendix)

$$i\left[\phi^*\frac{\partial\phi}{\partial t} - \frac{\partial\phi^*}{\partial t}\phi\right]. \qquad \ldots\ldots(2.27)$$

This, however, is unsatisfactory. Whereas $\psi^*\psi$ is a positive definite quantity (as probability density should be), there is no reason whatever why expression (2.27) should be positive. It seems impossible to find, in relativistic particle mechanics, an expression for probability density with even the elementary property of being positive.

This objection to the wave equation (2.26) was noted in the early days of quantum theory. It held up all progress in the subject, and for a long time it seemed that we would never have a satisfactory equation for a particle with relativistic energy.

In 1934, for the first time, Pauli and Weisskopf noticed the great similarity between equations (2.26) and (2.3). In a sense, the electromagnetic equation is a special case of the equation (2.26) for $m=0$.

Once this similarity was noticed, the solution of the difficulties was obvious. ϕ must not be treated as a wave function, describing a particular particle. This, like $\mathbf{A}(\mathbf{r}, t)$, is an operator, describing an entire assembly, a field of relativistic particles, each with an intrinsic rest mass m.

The theory of the matter field now follows exactly the same pattern as the theory of the radiation field. A commutation relationship is postulated between ϕ and $d\phi/dt$. ϕ is Fourier analysed, the Fourier amplitudes acting as creation and annihilation operators. There are just two essential differences from the case of the Maxwell field.

Firstly the field is described by a scalar ϕ, rather than by a vector quantity $\mathbf{A}$. The number of the independent components of the field operator (1 in the case of ϕ, and 3 in the case of $\mathbf{A}$), gives the intrinsic spin of the particle described by the field. Thus the ' quanta ' of the ϕ field are spin zero particles ($2s+1=1$), rather

than particles of spin one (photons). Secondly ϕ is a complex operator, as against **A**, which is real (Hermitian). Setting $\phi=\phi_1+i\phi_2$, it is clear that the field ϕ carries two intrinsic degrees of freedom. These we associate with the two degrees of freedom for charge. The theory, thus, describes ' scalar ' particles of positive or negative charge.

We can, in fact, construct an operator which has the meaning of total charge in the field. Let

$$Q=ie\int\left(\phi^*\frac{\partial\phi}{\partial t}-\frac{\partial\phi^*}{\partial t}\phi\right)d^3x. \qquad \ldots\ldots(2.28)$$

The integrand is the same expression which we tried to interpret as probability density. Using the Fourier expansion of ϕ and ϕ^*, it can be shown that the eigenvalues of Q are numbers ne, where n's are positive as well as negative integers. The eigenvalue equation

$$Q\Psi=ne\Psi \qquad \ldots\ldots(2.29)$$

has a simple interpretation—it tells us that there are just n units (+ or −) of charge in the field.

It is also easy to verify that

$$Q\phi^*=\phi^*(Q+e). \qquad \ldots\ldots(2.30)$$

Using (2.29) we see that

$$Q(\phi^*\Psi)=(n+1)e(\phi^*\Psi), \qquad \ldots\ldots(2.31)$$

implying that the system in a state described by $\phi^*\Psi$ contains one unit more of positive charge than in the state Ψ. ϕ^* is therefore a creation operator—a creation operator for positive charge. This it may accomplish of course, either by creating a positively charged particle, or by annihilating a negatively charged one.

On the formal side, the development for the matter field differs but little from the radiation field case. The important point is that the requirement of the special relativity principle forces us to abandon all attempts at constructing one-particle wave equations. The logic of the mathematics leads inexorably to a many-particle picture—the picture of a field, of an assembly of particles, described by the field operator $\phi(\mathbf{r}, t)$.

§ 5. The Dirac Equation

We have here reversed the historical order. In actual fact, even before the work of Pauli and Weisskopf, Dirac (1928) had almost succeeded in constructing a wave equation for a single relativistic particle. Dirac postulated that the behaviour of such a particle should be described by the wave equation

$$\left(\gamma_1\frac{\partial}{\partial x}+\gamma_2\frac{\partial}{\partial y}+\gamma_3\frac{\partial}{\partial z}-\gamma_0\frac{\partial}{c\partial t}+\frac{mc}{\hbar}\right)\psi(\mathbf{r}, t)=0. \qquad \ldots\ldots(2.32)$$

Here ψ is 4×1 matrix (spinor) and γ_1, γ_2, γ_3, γ_0, are 4×4 matrices. One convenient choice for the γ's is the following

$$\gamma_1=\begin{pmatrix}0&0&0&-i\\0&0&-i&0\\0&i&0&0\\i&0&0&0\end{pmatrix}, \quad \gamma_2=\begin{pmatrix}0&0&0&-1\\0&0&+1&0\\0&+1&0&0\\-1&0&0&0\end{pmatrix}, \quad \ldots\ldots(2.33)$$

$$\gamma_3=\begin{pmatrix}0 & 0 & -i & 0\\ 0 & 0 & 0 & +i\\ i & 0 & 0 & 0\\ 0 & -i & 0 & 0\end{pmatrix}, \qquad \gamma_0=\begin{pmatrix}i & 0 & 0 & 0\\ 0 & i & 0 & 0\\ 0 & 0 & -i & 0\\ 0 & 0 & 0 & -i\end{pmatrix}. \quad \ldots\ldots(2.33)$$

Dirac was able to show that if ψ is interpreted as a wave function, the corresponding particle has the following properties:

(i) The energy and momentum of the particle are connected by the relativistic relation $E^2=m^2c^4+p^2c^2$. (ii) The particle has spin $\frac{1}{2}\hbar$ and a magnetic moment of one Bohr magneton (thus explaining the magnetic anomaly of the spin). (iii) $\rho=\psi^*\psi$ gives the probability of its being found at space point $\mathbf{r}$, at time t, and has the essential property of being a positive quantity.

All this was extremely satisfactory. Dirac had completely circumvented the negative probability difficulties which had wrecked the previous attempts at formulating relativistic wave equations. This, however, was accomplished at a very heavy price; to avoid negative probabilities Dirac was forced to introduce negative energies into his theory. Specifically, it was found that the energy values for the particle described by ψ could be $-\sqrt{(m^2c^4+p^2c^2)}$ just as well as $+\sqrt{(m^2c^4+p^2c^2)}$ and by no means could these negative energies be excluded.

The ingenious solution which Dirac proposed for overcoming these difficulties is well known. Taking advantage of the fact that spin $\frac{1}{2}$ particles obey Pauli's exclusion principle, Dirac postulated the existence of an infinite number of electrons filling all possible negative energy states. A hole in this ' sea ' of (otherwise unobservable) electrons would then be interpreted as a positively charged particle with positive energy—the positron.

All this, of course, was needed to save the one-particle theory. To give a consistent meaning to a one-particle wave function one is forced, in this context to postulate the existence, besides this particle, of an infinite number!

Once again it is clear that the correct and the simplest way to proceed (see for example Pauli 1941) is to assume that $\psi(\mathbf{r}, t)$ is a field operator and not a wave function and that it describes a charge-carrying field, the field of electrons and positrons. $\psi(\mathbf{r}, t)$, just like $\phi(\mathbf{r}, t)$ annihilates charge; it may do so by annihilating electrons, or by creating positrons. The electron and the positron bear to each other a relation no more mysterious than that between positive and negative mesons. The total energy of the field is always positive; it is the sum of the energies of the electrons and the positrons present. The only important difference in the theory of the operator ψ from that outlined above for **A** is that we have to postulate an anticommutation relation in place of (2.9). Thus we assume

$$\psi(\mathbf{r}, t)\psi^*(\mathbf{r}', t)+\psi^*(\mathbf{r}', t)\psi(\mathbf{r}, t)=\delta(\mathbf{r}-\mathbf{r}'), \qquad \ldots\ldots(2.34)$$

rather than

$$\psi(\mathbf{r}, t)\psi^*(\mathbf{r}', t)-\psi^*(\mathbf{r}', t)\psi(\mathbf{r}, t)=\delta(\mathbf{r}-\mathbf{r}').$$

This corresponds to the use of Fermi, rather than Bose statistics for spin $\frac{1}{2}$ particles. The state of the system is completely described by the eigenvalues of the energy and the total charge operators.

The latter operator is simply the space integral of the expression which Dirac postulated for probability density in his one-particle theory. Thus charge density $\rho=ie\psi^*\psi$ (2.33), and the corresponding expression for current density j_i is

$$ie\bar{\psi}\gamma_i\psi \qquad \ldots\ldots(2.35)$$

where $\bar{\psi}=i\psi^*\gamma_0$. ρ and j_i form a four-vector in the usual manner, which may be written

$$j_\mu = ie\bar{\psi}\gamma_\mu\psi \text{ with } \mu = 0, 1, 2, 3. \qquad \ldots\ldots(2.36)$$

§ 6. General Field Theory

The procedure described above—that of treating the field function as an operator, postulating a commutation relation and Fourier analysing this operator to obtain creation and annihilation operators—is the standard procedure through which field theory works. The particles described by each type of field possess a number of intrinsic properties. These are a reflection of the various transformation properties of the field functions themselves.

As remarked before, if the field function $\phi(x)$, $x=(\mathbf{r}, t)$, is a scalar, the 'quanta' of the field have zero spin. If this function is a four-vector, $\phi_\mu(x)$, $\mu=0, 1, 2, 3$, it can be shown quite generally that only three of these four components are independent. These three components can then be associated with the three spin states of a particle of intrinsic spin unity. The quanta of fields satisfying Dirac's equation, on the other hand, possess an intrinsic spin $\frac{1}{2}\hbar$. The intrinsic parity of the particles is determined by the transformation properties of the field operator for space reflection. If under a transformation $\mathbf{r}\to-\mathbf{r}$, $t\to t$, $\phi(\mathbf{r}, t)$ transforms to $\phi'(-\mathbf{r}, t)$ where $\phi'(-\mathbf{r}, t)=\phi(-\mathbf{r}, t)$ the field is called 'scalar'. If, however, $\phi'(-\mathbf{r}, t)=-\phi(-\mathbf{r}, t)$ the field is called 'pseudoscalar' and the particles it describes have an intrinsic parity -1. Pseudoscalar fields were first considered by Kemmer (1938).

APPENDIX

In the text, we have remarked that for wave equation (2.5), the only possible choice for ρ is

$$i\left(\phi^*\frac{\partial\phi}{\partial t}-\frac{\partial\phi^*}{\partial t}\phi\right).$$

This is for the following reason: in order to conserve total probability ρ the probability density must satisfy a continuity equation of the form

$$\frac{\partial\rho}{\partial t}+\operatorname{div}\mathbf{j}=0. \qquad \ldots\ldots(A1)$$

Consider the equations

$$\nabla^2\phi-\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\phi=\frac{m^2c^2}{\hbar^2}\phi, \qquad \ldots\ldots(A2)$$

$$\nabla^2\phi^*-\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\phi^*=\frac{m^2c^2}{\hbar^2}\phi. \qquad \ldots\ldots(A3)$$

Multiply (A2) by ϕ^* and (A3) by ϕ, and subtract. We obtain:

$$\operatorname{div}(\phi^*\operatorname{grad}\phi-\phi\operatorname{grad}\phi^*)+\frac{i}{c}\frac{\partial}{\partial t}\left(\phi\frac{\partial\phi^*}{\partial t}-\frac{\partial\phi}{\partial t}\phi^*\right)=0. \qquad \ldots\ldots(A4)$$

This equation has the required form (and is the only one of this type) justifying our choice of ρ.

Part III. INTERACTIONS OF QUANTIZED FIELDS

By P. T. Matthews

§ 1. Annihilation and Creation Operators

The essential point of Part II is that a quantized field may be considered as an assembly of identical particles. These may be neutral, or may be positively and negatively charged, according to whether the field is real or complex. It is usual to classify them according to their momentum, which, for convenience, we shall assume to take certain discrete values $\mathbf{k}_1, \mathbf{k}_2, \ldots, \mathbf{k}_i, \ldots$ ($\mathbf{k}$ is the $\hbar\omega/c$ of Part II: see equation (2.13)). A state of the assembly is then determined by giving the number of particles, $n_1, n_2, \ldots, n_i, \ldots$, in each momentum state—the occupation numbers as they are called. The state function—the analogue, in this picture, of the wave function—is then $\Psi(n_1, n_2, \ldots, n_i, \ldots)$.

We now want to consider the interactions of such systems. Since the only observable property of the system is the occupation numbers, the only observable change in the system is a change in these numbers. At first sight this seems to bear little relation to what we usually mean by an interaction or the reactions which it may produce. We will now show that it provides a slightly unusual but, in fact, perfectly adequate description of all the familiar reactions which can take place.

Suppose the system consists of a number of electrons in a box (we can forget the positrons for the moment). Suppose the interaction is that of an electron bouncing against the sides of the box, as illustrated in figure 3.1. This is described in our formalism as a change in occupation numbers—the annihilation of an electron in the state $\mathbf{k}_1$ and the creation of an electron in the state $\mathbf{k}_2$. Since the electrons are indistinguishable, it does not matter whether one thinks of it as the 'same' or a 'new' electron which gets created.

If the system consists of an assembly of electrons and photons in interaction, then the place of the box in this transition is taken by a photon, which gets created as the electron is shifted from one momentum state to another.

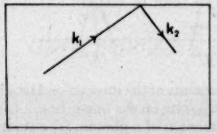


Figure 3.1. A particle bouncing against the side of a box is described as the annihilation of a particle in state $\mathbf{k}_1$ and the creation of a particle in state $\mathbf{k}_2$.

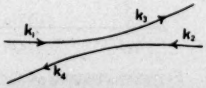


Figure 3.2. Typical orbits in the classical scattering of two electrons by each other. The corresponding process is described in field theory as the annihilation of electrons in states $\mathbf{k}_1$ and $\mathbf{k}_2$, and the creation of electrons in states $\mathbf{k}_3$ and $\mathbf{k}_4$.

If two electrons scatter against each other this would be described classically by the two orbits shown in figure 3.2. These can be calculated exactly in classical theory. In our formalism this is a transition from an assembly, in which the $\mathbf{k}_1$ and $\mathbf{k}_2$ states are occupied, to one in which the $\mathbf{k}_3$ and $\mathbf{k}_4$ states are occupied. We

also describe it as the annihilation of the electrons in states $\mathbf{k}_1$ and $\mathbf{k}_2$ and the creation of electrons in states $\mathbf{k}_3$ and $\mathbf{k}_4$. We wish to calculate the probability of such an event taking place.

These transitions are most simply described in terms of annihilation and creation operators, which have the effect of altering the occupation numbers of the system. This is very satisfactory because the second main point of the previous Part was that these operators are just the fields $\psi(x)$, $A(x)$, etc., from which we started.

As always in quantum theory, the interaction is included by adding to the energy, or Hamiltonian, of the system a term describing the interaction energy, which, in our case, will be a function of the field operators.

§ 2. Electrodynamics

Let us concentrate now on electrodynamics, the interaction of an electron–positron assembly with a photon assembly. We can easily deduce the interaction energy from classical physics.

If a particle of charge e is in a potential $V(x)$ the energy is $eV(x)$. If the charge is spread over a region with density $\rho(x)$ the energy density is $\rho(x)V(x)$ or, $(A_0(x)=V(x))$, $j_0(x)A_0(x)$ in the relativistic notation of (2.36). The whole interaction is a scalar† and we have clearly got just one component of the scalar product of two vectors. The complete expression for the interaction energy is thus

$$H(x)=j_\mu(x)A_\mu(x). \qquad \ldots\ldots(3.1)$$

In terms of the electron–positron field operators this is, by (2.36),

$$H(x)=ie\bar{\psi}(x)\gamma_\mu\psi(x)A_\mu(x). \qquad \ldots\ldots(3.2)$$

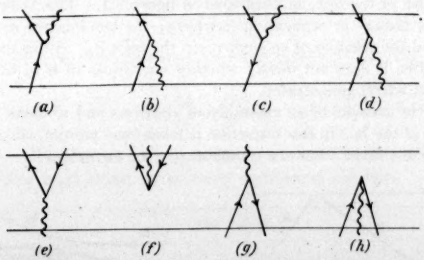


Figure 3.3. Feynman diagrams of the matrix elements of the interaction Hamiltonian (3.2). The initial state is on the lower line, the final state on the upper line. (*a*) an electron going from one state to another with the emission of a photon, (*c*) a similar transition for a positron.

Now let us consider the changes in occupation numbers which can be produced directly by this interaction. A may either create or annihilate a photon. $\bar{\psi}$ creates electrons or annihilates positrons, ψ annihilates electrons or creates positrons. Combining these in all possible ways, one sees that the whole operator has matrix elements between the states depicted in figure 3.3. The initial state is that on the lower line. Thus (*a*) denotes an initial state with just one electron of specified

† Strictly speaking the Hamiltonian is the zero component of a four-vector (like ρ_0 or A_0). The Lagrangian is a true scalar and the Hamiltonian is constructed from it.

momentum, represented by the arrow. This electron is annihilated by ψ, an electron in a different momentum state is created by $\bar{\psi}$. At the same time a photon is created by A, as illustrated by the wavy line. The final state of the assembly is thus an electron with some different momentum and a photon also of some specified momentum. (*b*) is a similar transition except that here there is a photon in the initial state, which is annihilated by A, leading to a final state with just one electron. Diagrams (*c*) and (*d*) denote the same processes, only the electron is replaced by a positron. We denote this by arrows pointing from the final to the initial states. The other graphs are then self-explanatory, (*e*) for example representing the annihilation of an initial photon by A and the creation in the final state of an electron–positron pair. There may of course be other particles, electrons or photons, in the assembly, which are not affected by the transition, and just carry through from the initial to the final state. We have not bothered to include these in the diagrams.

The arrow notation for electrons and positrons, first introduced by Feynman (1949), is the crucial step in the development of his graphical representation of events. The result is that, although we have constructed eight different transitions, there is essentially only one diagram, consisting of a directed line running through the interaction point, and one wavy line. It is easily seen that the fact that the arrows run through the interaction point, and are neither both directed towards it, nor both away from it, expresses the conservation of charge.

None of the transitions which we have drawn can actually take place as a real process, because for this to happen both energy and momentum must be conserved. It is obvious that in (*f*) and (*g*) energy alone cannot even be conserved. A simple calculation shows that it is not possible to satisfy both energy and momentum conservation for any of the other transitions. However, as was shown in Part I, energy need not be conserved in virtual transitions, and these transitions may thus combine together as virtual steps to form a possible real process.

The actual consequences of the interaction (3.2) may be calculated by the use of time dependent perturbation theory. This states that the probability per unit time of a transition from an initial state Ψ_i to a final state Ψ_f is

$$\omega = \frac{2\pi}{\hbar}|\langle f|M|i\rangle|^2\rho(E), \qquad \ldots\ldots(3.3)$$

where $\rho(E)$ is the density of final states per unit energy range, and

$$\langle f|M|i\rangle = \langle f|H|i\rangle + \sum_n \frac{\langle f|H|n\rangle\langle n|H|i\rangle}{E_n - E} \qquad \ldots\ldots(3.4)$$

$$+ \sum_n \sum_m \frac{\langle f|H|n\rangle\langle n|H|m\rangle\langle m|H|i\rangle}{(E_n - E)\,(E_m - E)} + \ldots$$

H is the interaction energy and the sums over n and m are taken over all possible states. The perturbation theory in this form is very awkward for our purpose. We have been at great lengths to set up a relativistically invariant theory. While the results of using the formula (3.4) are in fact relativistically invariant, (3.4) itself does not have a covariant form. It brings about a great simplification if one can preserve the covariant form of the theory explicitly at each stage in the calculation. This has been achieved in a reformulation of perturbation theory due to Feynman (1948, 1949) and Dyson (1949) which enormously simplifies the calculation, particularly of higher terms in the perturbation series.

As far as we are concerned, all that matters is that, in both formulations, the nth term of the perturbation series depends on the matrix product of the interaction Hamiltonian, taken n times. The resulting contribution to a particular scattering process can be represented by diagrams made up of n vertices of the type depicted in figure 3.3, combined together to give the required particles in the initial and final states, as specified by Ψ_i and Ψ_f.

§ 3. Scattering Phenomena

In this way we can describe all the main processes of electrodynamics. Consider for example the scattering of electrons. In the lowest order this is represented by the graphs of figure 3.4. Each vertex is of the type shown in figure 3.3, and the

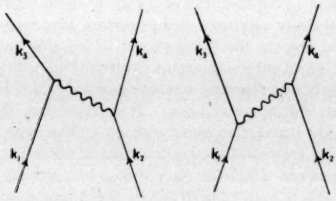


Figure 3.4. Feynman diagrams for the Møller (Coulomb) interaction.

whole interaction may be described as the exchange of a virtual photon. As described in Part I, this is precisely equivalent to the Coulomb interaction in the non-relativistic limit. Our theory is relativistic so that the extra terms of the Møller interaction are also included.

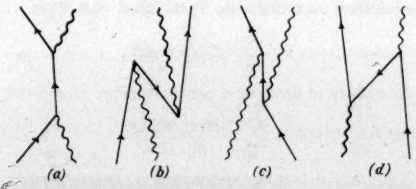


Figure 3.5. Feynman diagrams for the Compton effect. (*b*) and (*d*) involve virtual pair creation.

Now consider the Compton effect, or scattering of photons by free electrons. The possible graphs of lowest order are shown in figure 3.5. Graphs (*b*) and (*d*) involve the creation of pairs in the intermediate states. In the Feynman–Dyson form of perturbation theory, these graphs are not merely convenient ways of visualizing what is going on. There is a direct one-to-one relation between the diagrams and the mathematical theory, and there are some very simple rules which enable one to write down, from the graphs, the integrals which then determine the matrix element $\langle f|M|i\rangle$ for the process. An important point of these rules is that

it is only necessary to take into account explicitly graphs such as (*a*) and (*c*), which are essentially different in their structure. Graphs such as (*a*) and (*b*), (or (*c*) and (*d*)), which are topologically identical and differ only in the order in time of the different vertices, are treated as a single graph. The same too, applies to the two topologically identical graphs of figure 3.4. In future we will consider only one typical graph for each process.

The other main processes of electrodynamics, for which atomic binding may be neglected, are all obtained from the Compton diagrams by a reinterpretation of the external lines. The Coulomb field of a heavy nucleus may be considered as a source of longitudinal photons and is represented diagrammatically by a wavy line

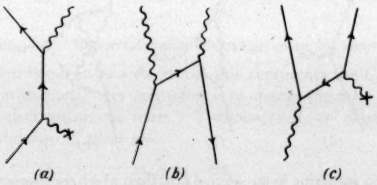


Figure 3.6. Feynman diagrams for (*a*) bremsstrahlung, (*b*) two-photon pair annihilation, (*c*) pair creation by γ-rays in the presence of a nucleus.

ending in a cross. Figure 3.6(*a*) is thus bremsstrahlung (the scattering of an electron in the field of a nucleus with the emission of a real photon); figure 6(*b*) is the annihilation of an electron–positron pair into two photons; (*c*) is the creation of a pair by a photon in the field of a nucleus. Processes for which the atomic binding of the electron may not be neglected are more complicated in practice but not in principle.

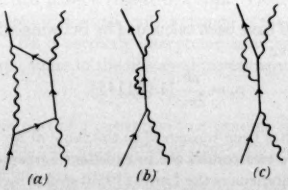


Figure 3.7. Radiative corrections to the Compton effect.

So far only the lowest order term in the perturbation series, or the corresponding simplest diagrams, have been considered. The next order terms will involve the interaction of the Hamiltonian two more times, and are represented by the inclusion of an extra virtual photon in the lowest order diagrams. Some of the corrections to figure 3.5(*a*) are shown in figure 3.7. The expansion is in powers of the interaction energy or the charge e. This constant determines the strength of the interaction. Since only even powers occur the expansion is in powers of e^2, or strictly

$e^2/4\pi\hbar c$, which is a small number very nearly equal to 1/137 (the fine structure constant). The corrections are thus all small, and of no great interest, as they give effects which are masked by the experimental errors.

§ 4. Anomalous Electron Moment and the Lamb Shift

There are, however, several effects for which the higher order corrections can actually be observed. Of these the simplest is the anomalous magnetic moment of the electron. If an electron is scattered in an external field, to lowest order in e, the process is represented by figure 3.8(a). This is now a possible real transition,

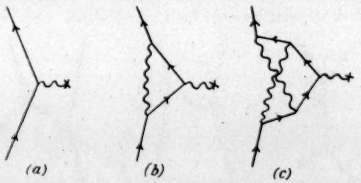


Figure 3.8. Electron scattering in an external field from which the magnetic moment may be deduced: (a) first Born approximation giving one Bohr magneton; (b) radiative corrections of order e^3; (c) typical correction of order e^4.

since momentum need not be conserved in scattering in an external field. If the external field is a pure magnetic field, $\mathbf{H}$, the scattering can be shown to correspond to an interaction energy of $(e\hbar/2mc)\boldsymbol{\sigma}.\mathbf{H}$, showing that the electron has, to this order in e, a magnetic moment of one Bohr magneton. However, there are clearly corrections, those of order e^2 given by 3.8(b), and some of order e^4 coming from graphs such as 3.8(c). Thus the true magnetic moment is

$$\mu_e = \frac{e\hbar}{2mc}[1 + Ae^2 + Be^4 + \ldots]. \qquad \ldots\ldots(3.6)$$

The coefficients A and B have been calculated by Schwinger (1949) and by Karplus and Kroll (1950) and give

$$\mu_e = \frac{e\hbar}{2mc}[1{\cdot}001145]. \qquad \ldots\ldots(3.7)$$

The magnetic moment has been measured to the same order of accuracy by Kusch and others (1951). The two results are in excellent agreement.

Another similar phenomena is the Lamb (1950) shift which is the splitting of the $2S_{1/2}$ and $2P_{1/2}$ levels in hydrogen, due to the emission and reabsorption of virtual photons by the electron in the bound state. The same type of effect has been observed in positronium and μ-mesic atoms. For all these, agreement between theory and experiment is excellent. For the anomalous electron moment the agreement is to one part in a million.

§ 5. Renormalization

In all these calculations great care has to be taken to express the results in terms of the experimental mass and charge of the electron. The parameter m which

appeared in equation (2.32) (Part II) is the mechanical mass. It is often called the ' bare ' mass. A free electron, in our formalism, may emit a photon and subsequently reabsorb it (figure 3.9(*a*)). Just as figure 3.4 represents the relativistic

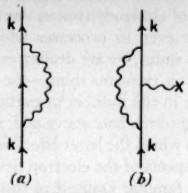


Figure 3.9. Renormalization effects: (*a*) mass, (*b*) charge.

Coulomb interaction between two electrons, this represents the Coulomb interaction of an electron with itself. Its only effect is to change the mass of the particle—to add δm to the electromagnetic mass. The observed, or experimental, mass, m_{ex}, is, of course the sum of these two

$$m_{ex}=m+\delta m. \qquad \ldots\ldots(3.8)$$

When the emission and absorption, by a virtual electron, of a photon takes place in the middle of a more complicated interaction (see for example figure 3.7(*b*)), its main effect is again to change the electron mass from the mechanical to its observed experimental value. In this case there is, however, an additional physically significant correction to the process under consideration, which shows up clearly after the whole expression has been written in terms of the observed mass.

A very similar effect occurs with the charge. The parameter e which we first introduce is the ' bare ' charge. An electron may suffer a formal scattering in an external field, in which it undergoes no change of momentum (neither of magnitude nor of direction). This formal interaction may also be ' straddled ' by an emission and reabsorption of a virtual photon (figure 3.9(*b*)). There is no contribution to the observed scattering from this graph, only a change in the energy of a particle in a constant static field. This is correctly interpreted as a change δe† in the charge to bring it from its ' bare ' value to the observed experimental value,

$$e_{ex}=e+\delta e. \qquad \ldots\ldots(3.9)$$

The physically significant part of a correction to a genuine scattering, such as that shown in figure 3.8(*b*) is that part which remains after everything has been expressed in terms of e_{ex} rather than e.

The elimination of the ' bare ' mass and charge in favour of their observed values is known as mass and charge renormalization. It so happens that the expressions for δm and δe, if one wishes to evaluate them, are both infinite. This is of no physical significance, because these quantities are unobservable, and do not appear in the final results for cross sections, level shifts, or anomalous magnetic moments. All these latter are given, by the theory, in terms of the observed mass

† There are other effects, which also contribute to δe, which have been omitted so as not to obscure the main idea. For a more detailed account in non-mathematical language see Matthews and Salam (1951). This also links up with the more physical interpretation of charge renormalization given in Part I.

and charge, which are of course finite. The important point is that, after mass and charge renormalization has been carried out, all the physical consequences of the theory are perfectly well defined, finite, and, in fact, in very detailed agreement with experiment (Schwinger 1949, Dyson 1949).

Before leaving the subject of electrodynamics, one more point should be made. It is important to realize that even in processes which involve only one or two particles in the initial and final states, we are dealing essentially with the interaction of particle *assemblies* or—which is the same thing—the interaction of two quantized operator fields. This increase in the number of particles comes from the possible creation of virtual photons in intermediate states and, most especially, from virtual pair creation. In a problem in which the latter effect may consistently be neglected, one may ignore the operator aspects of the electron field and treat Dirac's equation as a one-particle equation. A familiar example of this is in the theory of the scattering of an electron in an external static field (ignoring radiative corrections). However the detailed agreement with experiment in the Lamb shift and similar phenomena depends essentially on the field, or particle assembly concept, and could not possibly be obtained from the single-particle point of view. (For example the reader may be interested to check that by ordering the vertices in time, as is done in figure 3.5(*b*), figure 3.8(*c*) can be shown to contain a contribution from an assembly in which there are three electrons, two positrons and three photons present).

§ 6. Meson Theories

We now pass to the theory of meson–nucleon interactions. This is developed in close analogy with electrodynamics, the electron–positron field being replaced by a nucleon field (we need not distinguish at this stage between protons and neutrons), and the electromagnetic potential $A(x)$ giving place to $\phi(x)$ the meson field. It is a fundamental assumption of the theory that anti-nucleons exist and should be created, when enough energy is available. It will be a fundamental blow to all current theories if this proves not to be the case.

The interaction energy may take a variety of detailed forms to which we return below. As far as the operators are concerned it has the same shape as the electromagnetic interaction, the charge e being replaced by some nuclear coupling constant g. Thus

$$H \simeq g\bar{\psi}\psi\phi. \qquad \ldots\ldots(3.10)$$

The resulting changes in occupation numbers, brought about directly by the interaction Hamiltonian, are thus exactly the same as those depicted in figure 3.3, with the appropriate change in interpretation from electrons to nucleons, and photons to mesons. As before, these transitions cannot take place as real processes, since energy and momentum cannot be conserved.

The real processes, which necessarily follow from the assumption of such an interaction, are depicted in Feynman's graphical language in figure 3.10, in the lowest order in which they appear in an expansion in powers of g; mesons are denoted by dotted, and photons by wavy lines. Only one typical graph for each process is drawn. (*a*) is the effect for which the mesons were postulated by Yukawa (1935) in the first place, namely the scattering of nucleons on nucleons. Just as the exchange of a virtual photon gives rise, in the non-relativistic limit, to an effective potential of the Coulomb form $1/r$, so this exchange of a virtual meson produces, in the same limit, the Yukawa potential $e^{-\kappa r}/r$ $(\kappa = m_\pi c/\hbar)$.

The process (*b*) is the creation of a real meson in the collision of two nucleons ($N+N \rightarrow N+N+\pi$). This must take place as soon as there is sufficient kinetic energy among the initial particles to produce the rest energy of the π-meson in the final state. As the initial energy increases the theory implies that double and triple π-production will also set in.

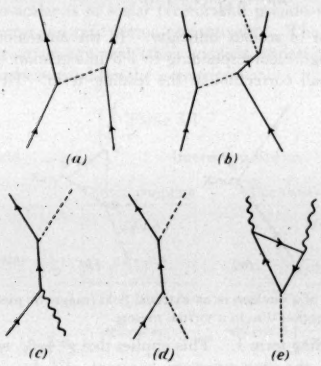


Figure 3.10. Typical diagrams of meson–nucleon interactions. Full lines denote nucleons, dotted lines mesons, and wavy lines photons. (*a*) Nucleon–nucleon scattering (Yukawa potential). (*b*) Meson production in a nucleon–nucleon collision. (*c*) Photo-meson production. (*d*) Meson–nucleon scattering. (*e*) Decay of neutral meson into two γ-rays.

In (*c*) the photo-production reaction is illustrated, ($\gamma+N \rightarrow N+\pi$). If photons of sufficient energy are incident on protons, then again π-mesons should be created.

Meson–nucleon scattering, depicted in (*d*), ($\pi+N \rightarrow N+\pi$), is the mesic analogue of the Compton effect.

(*e*) shows the decay of a neutral π-meson into two photons ($\pi^0 \rightarrow 2\gamma$). It is of interest that this reaction goes through the creation in a virtual state of a nucleon–anti-nucleon pair. Note that this mode of decay is only available to neutral mesons, since the photons cannot carry off any charge.

One more effect, which is implied by the interaction which we have assumed, is the anomalous magnetic moments of the nucleons. If a proton is scattered in an external magnetic field $\mathbf{H}$ it can be shown, just as for an electron, that the effective interaction energy is of the form

$$\frac{e\hbar}{2Mc}\,\boldsymbol{\sigma}.\,\mathbf{H}.$$

The proton should thus have a magnetic moment of one Bohr magneton (modified, of course, by the replacement of the electron by the proton mass), figure 3.11(*a*). In addition to the corrections due to virtual photons, one should expect (larger) corrections due to virtual meson exchanged as shown in figure 3.11(*b*). The proton

moment is thus given by an expansion

$$\mu_P = \frac{e\hbar}{2Mc}(1+Ag^2+Bg^4+\ldots). \qquad \ldots\ldots(3.11)$$

Now it is known from experiment that, roughly,

$$\mu_P = \frac{e\hbar}{2Mc}(1+1{\cdot}8). \qquad \ldots\ldots(3.12)$$

This is the first sign of serious difficulty. In the anomalous electron moment, equation (3.7), the figure corresponding to 1·8 in equation (3.12) was 0·001145, and really was a small correction to the leading term. Here we have 1·8 as a

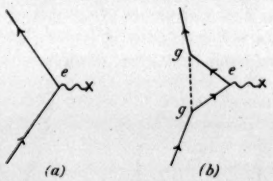


Figure 3.11. Scattering of a nucleon in an external field (magnetic moment). (*a*) First order in e. (*b*) A correction due to a virtual meson.

'correction' to the leading term 1. This implies that $g^2/4\pi\hbar c$, which is the analogue in meson theory of the fine-structure constant $e^2/4\pi\hbar c$, instead of being a small number like 1/137, is at least of the order of 1. The result is that our standard method of calculating effects as an expansion in g^2 is not going to be applicable. Although we may draw graphs for the processes, as we have done in figure 3.9, these are no longer in a simple one-to-one relation with the mathematics which correctly describes the processes. The interaction between two nucleons is not even approximately described by the exchange of a single virtual meson, but by a chaotic jumble of virtual mesons and, possibly, nucleon–anti-nucleon pairs. Similarly the free nucleon cannot be thought of as occasionally emitting and reabsorbing a virtual meson (as in the case of electrons, and photons, figure 3.9(*a*)), but rather as a 'bare' nucleon continuously surrounded by an appreciable 'cloud' of virtual mesons and, possibly, pairs, which give rise among other things to the large anomalous magnetic moments.

§ 7. Possible Interactions

The qualitative picture which we have built up of the consequences of the Yukawa hypothesis is independent of the details of the interaction which we assume. The possible forms which the theory can take were deduced by Kemmer (1938), in a fundamental paper which laid the foundations of the whole theory. These are given in table 3.1.

The main condition which the interaction energy must satisfy is that it is a scalar. If the meson field is itself a scalar, it must be multiplied by a scalar combination of the nucleon field operators, giving the so-called direct coupling. The interaction may also be constructed from the derivative $\partial\phi/\partial x_\mu$ which is a vector and must thus be multiplied by a vector, namely the nucleon current. This gives the so-called derivative coupling. If the meson field is a pseudo-scalar, then it must be multiplied

by a pseudo-scalar combination of the nucleon fields. This proves to be $\bar{\psi}\gamma_5\psi$ where

$$\gamma_5 = \gamma_0\gamma_1\gamma_2\gamma_3. \qquad \ldots\ldots(3.13)$$

γ_5 is thus a new, 4×4, Dirac matrix which is easily constructed from those already introduced (equation (2.33)). The general principle for constructing the interactions should now be apparent. The pseudo-vector field is in the same relation to the vector, as pseudo-scalar is to scalar (vector and pseudo-vector are sometimes referred to as polar vector and axial vector, respectively). The factors i are included to make the interaction energy real (their explicit appearance depends on the convention which has been used for the γ-matrices).

Table 3.1

Meson Field	Interaction Energy	
	Direct coupling	Derivative coupling
Scalar	$g\bar{\psi}\psi\phi$	$ig\bar{\psi}\gamma_\mu\psi\dfrac{\partial\phi}{\partial x_\mu}$
Pseudo-scalar	$g\bar{\psi}\gamma_5\psi\phi$	$g\bar{\psi}\gamma_\mu\gamma_5\psi\dfrac{\partial\phi}{\partial x_\mu}$
Vector	$ig\bar{\psi}\gamma_\mu\psi\phi_\mu$	$ig\bar{\psi}\gamma_\mu\gamma_\nu\psi\left(\dfrac{\partial\phi_\mu}{\partial x_\nu}-\dfrac{\partial\phi_\nu}{\partial x_\mu}\right)$
Pseudo-vector	$g\bar{\psi}\gamma_\mu\gamma_5\psi\phi_\mu$	$g\bar{\psi}\gamma_\mu\gamma_\nu\gamma_5\psi\left(\dfrac{\partial\phi_\mu}{\partial x_\nu}-\dfrac{\partial\phi_\nu}{\partial x_\mu}\right)$

If we assume for the moment that perturbation theory calculations are valid for all these interactions, it is again, of course, important to express the cross sections etc., in terms of the observed, and not the ' bare ', masses and coupling constants g. That is to say, the masses—the meson mass also in this case—and coupling constants must be renormalized.

Now an important distinction arises. After renormalization the terms in the perturbation series are rendered finite and unambiguous only for the direct coupling of scalar and pseudo-scalar mesons. For all the other interactions involving charged mesons the predictions even after renormalization for any physical effect are all infinite (Matthews and Salam 1951). It may be that it will one day be shown how physically sensible results can be obtained from these ' unrenormalizable ' interactions, but in the meanwhile there is a very strong incentive for trying to explain all the physical facts in terms of interactions which give finite predictions after mass and coupling constant renormalization.

Part IV. PROPERTIES OF THE π-MESON

By P. T. Matthews

§ 1. π and μ Mesons

We have given in Part III the basis of meson theory. We have in fact got to the point in meson physics which, historically, had been reached at the outbreak of the war. In 1935 Yukawa made the very far reaching suggestion that the nuclear potential was due to the virtual exchange of mesons. By 1938 the possible forms which the theory could take had been enumerated by Kemmer, who pointed out, in particular, that the meson field might be pseudo-scalar. In terms of the particles, this means that a meson may have intrinsic negative parity. It was also known, from the general argument given in Part I, that to give the observed range of the nuclear potential the meson mass must be between 200 and 300 times the mass of the electron.

One important experimental development of this period has not yet been mentioned. Shortly after Yukawa's suggestion had been made, a meson of the required mass was observed by Anderson and Neddermeyer (1937). This was interpreted at first as dramatic evidence for Yukawa's hypothesis. However, this particle proved to be more trouble than it was worth, as subsequent investigation showed that it interacted very weakly with nucleons, and could not possibly play the role which Yukawa had predicted for it. This apparent paradox was cleared up in 1947 when Powell and the Bristol group (Lattes, Occhialini and Powell 1947), using the newly developed photographic techniques, showed that there are, in fact, two types of meson in the required mass range. These are now known as the π meson and the μ meson. Their masses and decay schemes are given in table 4.1. The μ meson has effectively no interaction with nucleons (weaker by a factor of 10^{12} than that required by Yukawa), and it was this relatively long-lived particle which was observed by Anderson and Neddermeyer (1937). The short-lived parent particle, π, is the one which had been predicted by Yukawa, and to which we will devote our attention exclusively from now on.

Table 4.1.

Meson	Mass	Decay scheme	Lifetime (sec)
$\pi^{\pm}$	273 (139)	$\pi^{\pm}\rightarrow\mu^{\pm}+\nu$	$2{\cdot}5\times10^{-8}$
$\mu^{\pm}$	207 (105)	$\mu^{\pm}\rightarrow e^{\pm}+2\nu$	$2{\cdot}1\times10^{-6}$
π^{0}	264 (135)	$\pi^{0}\rightarrow2\gamma$	$\sim10^{-14}$

Masses are given in units of the electron mass. Figures in brackets are mc^2 in MeV. The ν's appearing in the decay schemes are neutrinos.

The next big step forward was the construction in several places in America of particle accelerators of sufficient energy to create mesons artificially. The first to be completed was the proton synchro-cyclotron at Berkeley, California, giving a proton beam of about 340 MeV. This is above threshold for π meson production, and within a short time had provided a beautiful experimental confirmation of all the qualitative predictions of the Yukawa theory. All the processes sketched above were observed with relative cross sections of the order of magnitude to be expected.

Mesons were produced by nucleon–nucleon collisions and by photons on nucleons; π^0 mesons were observed to decay into photons (with a lifetime of about 10^{-14} sec) and, once meson beams had been produced, meson–nucleon scattering experiments were carried out. I want to stress very strongly this excellent qualitative agreement between the predictions of the theory and the experimental results. It tends to get obscured by the complete lack, so far, of any detailed quantitative agreement, which has been foreshadowed above.

We saw in the previous chapter that the same qualitative picture is given by a variety of different, formally possible, interactions. To begin to decide which is the meson–nucleon interaction actually observed in nature, one must determine not only the obvious properties, mass and charge, of the mesons, but also their intrinsic spin and their intrinsic parity. Experimental data accumulated rapidly after the completion of the Berkeley machine, and for a while it was hard to see the wood for the trees. I will make no attempt to review this material, but just pick out, in their logical order, the experimental results which give the information we require.

§ 2. Spin of π^0

We look first at the decay of the neutral π, or π^0. The existence of these particles was first established by Steinberger, Panofsky and Steller (1950), who observed the photons resulting from the impact of a photon beam on a nucleon. (The target was actually a nucleus—carbon or beryllium—but this does not concern us.) This is interpreted as a two-step process, first the photo-production of π^0's ($\gamma+\mathrm{N}\to\mathrm{N}+\pi^0$), and then the decay of the π^0's ($\pi^0\to\gamma+\gamma+(?)\gamma$). We are concerned only with the details of the second step. The essential experimental fact is that the decay is into two photons, and not three. This, as we shall show immediately, determines the spin of the π^0.

Consider the decay of the π^0 in its rest system. Suppose that the π^0 has intrinsic spin 1. Then the initial wave function must transform like the vector which determines the spin direction, just as the wave function of a photon transforms as the vector which gives its polarization. The final wave function of the two photons involves three vectors, $\boldsymbol{\epsilon}_1$ and $\boldsymbol{\epsilon}_2$ the two polarizations, and $\mathbf{k}$ the relative momentum. The wave function must be (i) bilinear in $\boldsymbol{\epsilon}_1$ and $\boldsymbol{\epsilon}_2$, and (ii) symmetric for an interchange of the two photons (that is, for interchange of the labels 1 and 2), since the photons obey Bose statistics; (iii) also $\boldsymbol{\epsilon}_1.\mathbf{k}=\boldsymbol{\epsilon}_2.\mathbf{k}=0$, since we are dealing here with real transverse photons.

The only bilinear vectors which we can construct are

(*a*) $\boldsymbol{\epsilon}_1\times\boldsymbol{\epsilon}_2$,

(*b*) $(\boldsymbol{\epsilon}_1.\boldsymbol{\epsilon}_2)\mathbf{k}$,

(*c*) $\mathbf{k}\times(\boldsymbol{\epsilon}_1\times\boldsymbol{\epsilon}_2)=\boldsymbol{\epsilon}_1(\mathbf{k}.\boldsymbol{\epsilon}_2)-\boldsymbol{\epsilon}_2(\mathbf{k}.\boldsymbol{\epsilon}_1)$.

(*a*) and (*b*) are excluded by (ii), since both change sign on the interchange of particles; (*c*) vanishes identically, by (iii). Thus the decay of a spin 1 particle into two photons is absolutely forbidden.

On the other hand, if the π^0 has intrinsic spin zero, the final wave function is then either a scalar or pseudo-scalar, according to the intrinsic parity of the π^0. Subject to the conditions (i), (ii) and (iii) we can construct two possible final wave functions,

(*e*) $\boldsymbol{\epsilon}_1.\boldsymbol{\epsilon}_2$,

(*f*) $(\boldsymbol{\epsilon}_1\times\boldsymbol{\epsilon}_2).\mathbf{k}$.

Thus this decay is perfectly possible and, if we exclude spins greater than one (which has been implicit in the discussion so far), the experiment definitely establishes zero spin for the π^0.

Notice that under a reflection of the space coordinates $\mathbf{k}$ will change sign. Thus (e) is a scalar, but (f) a pseudo-scalar. This shows that the decay photons of a scalar meson would have parallel polarization, whereas those from a pseudo-scalar meson would be perpendicularly polarized. This provides, in principle, a direct method of determining the intrinsic parity of the π^0.

§ 3. Spin of $\pi^{\pm}$

The next step is to establish the spin of the charged π's. Since these are manifestations of the same field, it is sufficient to consider either π^+ or π^-. If π^+ mesons are produced near threshold by the collision of protons on protons ($p+p \rightarrow \pi^+ + p + n$), the neutron and proton in the final state have little kinetic energy and there is a high probability of their coming off bound together to form a deuteron. The cross section for this process $p+p \rightarrow \pi^+ + D$, was observed by Cartwright, Richman, Whitehead and Wilcox (1951) using a 340 Mev proton beam, which corresponds to a 21 Mev meson energy in the centre-of-mass system. The equation for the cross section is

$$\sigma_{\text{prod}} = \frac{2\pi}{\hbar v_p} |M(E)|^2 \rho(\pi, \text{D}), \qquad \ldots\ldots(4.1)$$

where v_p is the velocity of the proton, $\rho(\pi, \text{D})$ is the density of the states per unit energy range of the π, D system. $M(E)$ is the matrix element for the process, which depends on the energy at which the experiment is performed. (The formula (4.1) differs from (3.3) by the factor v_p since we are working with cross sections and not probability per unit time.)

Given a meson beam, one may perform the reverse experiment using deuterons as a target. The cross section for this process, $\pi^+ + D \rightarrow p + p$, was observed by Clark, Roberts and Wilson (1951) and by Durbin, Loar and Steinberger (1951). The formulae for the cross section in a notation similar to that used above is

$$\sigma_{\text{abs}} = \frac{2\pi}{\hbar v_\pi} |M(E)|^2 \rho(\text{p}, \text{p}). \qquad \ldots\ldots(4.2)$$

By the principle of detailed balance the factors $M(E)$ are the same for both the π absorption and π production cross sections, provided both experiments are carried out at the same energy. The ratio is thus

$$\frac{\sigma_{\text{prod}}}{\sigma_{\text{abs}}} = \frac{v_\pi}{v_p} \frac{\rho(\pi, \text{D})}{\rho(\text{p}, \text{p})}. \qquad \ldots\ldots(4.3)$$

This follows independently of the details of the interaction. Now if the π meson has intrinsic spin s, $\rho(\pi, \text{D})$ will contain a factor $2s+1$, coming from the various directions in which this spin may be pointing in the final state. Since all the other factors in the above ratio can be calculated quite simply, quite inaccurate experiments are sufficient to determine whether this spin factor is 3 or 1. Both experiments show quite definitely that the factor is one, thus determining zero intrinsic spin for the charged π meson.

§ 4. Parity of $\pi^{\pm}$. The Panofsky Experiment

The only other experiment we wish to discuss is that performed by Panofsky *et al.* (1950) which, *inter alia*, determined the parity of π^-. In this experiment negative mesons of about 30 Mev were sent into hydrogen under pressure. It can be shown that these mesons will be slowed down, captured into the outer Bohr orbits of the hydrogen atoms, and cascade down to the ground state, by radiative transitions, in a matter of about 10^{-10} second. This is very short compared with the lifetime for $\pi \to \mu$ decay (see table 4.1). It is important to note that so far this is a purely electromagnetic effect. In the hydrogen ground state the meson, owing to its large mass, is almost ' inside ' the nucleus, and it is very quickly captured, due to the specific meson–nucleon interaction. According to our theory two processes may occur:

$$\pi^- + \mathrm{p} \to \mathrm{n} + \gamma \quad (a)$$

$$\pi^- + \mathrm{p} \to \mathrm{n} + \pi^0 \quad (b)$$

What Panofsky did was to observe the γ-rays coming off at a certain angle and plot their intensity against their energy. In (a) the photon will carry off nearly all the rest energy of the π^-. One would thus expect a sharp peak in the intensity plot at about 130 Mev. Its precise position determines the π^- mass. In process (b) the π^0 decays very rapidly into two photons of about 70 Mev each, giving rise to a broad hump in the photon intensity plot at about this energy. (The breadth of the hump is a Doppler broadening due to the fact that the π^0 does not decay at rest. Its maximum and minimum momenta in a given direction are determined by the $\pi^- - \pi^0$ mass difference. The width of the hump gives an accurate measurement of this difference, and thus fixes the π^0 mass.) The photon spectrum observed by Panofsky was just of the expected form which confirms beyond doubt that we are interpreting correctly what is going on.

One further point in this experiment is worth mentioning as it brings out a very general feature of photon–meson–nucleon interactions. One would expect the relative probabilities of reactions such as (a) and (b) to be roughly of the order of $e^2/g^2 (\simeq 1/137)$, since the one necessarily involves the weak electromagnetic interaction, whereas the other is a pure meso-nucleon effect. In this particular case, it was found that the two reactions were almost equally likely, indicating some kind of weak selection rule operating against the π^0 mode of decay.

However, the key result, as far as our present argument is concerned, comes from the companion experiment, in which the hydrogen was replaced by deuterium. The only difference is that the final π meson absorption process from the deuterium atomic ground state may now be expected to go in three ways:

$$\pi^- + \mathrm{D} \to \mathrm{n} + \mathrm{n} \quad (c)$$

$$\pi^- + \mathrm{D} \to \mathrm{n} + \mathrm{n} + \gamma \quad (d)$$

$$\pi^- + \mathrm{D} \to \mathrm{n} + \mathrm{n} + \pi^0 \quad (e)$$

The photons from (d) were observed. Those from the decay of the π^0 in (e) were not observed, showing that this effect is small, if not absolutely forbidden. By comparing the total intensity of photons with that in the hydrogen experiment, it was concluded that the reaction (c) definitely occurs. This is the crucial point.

Consider the final state in (c). We have two identical Fermi particles. The

wave function must thus be antisymmetric for an exchange of the two particles. Using the usual spectroscopic notation

$$^{2s+1}L_{j=L+s,\ \ldots,\ L-s},$$

the only possible states, with their corresponding parities, are

$$\begin{array}{cccc} ^1S_0, & ^3P_{0,\,1,\,2}, & ^1D_2, & \ldots \\ +, & - \ , & +, & \ldots \end{array}$$

The symmetric triplet spin states are combined with the odd angular momentum, the parity being $(-1)^L$. Note that all higher states will involve j values (total angular momentum) of at least two. In the initial state the deutron has spin one, its internal motion to a very good approximation has no angular momentum and the π^- is also absorbed from an atomic state of zero angular momentum. Thus the total angular momentum is $j=1$. Under space reflection the only change, if any, is from the intrinsic parity of the π^-. Now both parity and total angular momentum must be conserved in the transition. Since the only final state with $j=1$ has negative parity, and the reaction is definitely observed, the π^- must have intrinsic negative parity. The corresponding field is thus a pseudo-scalar. Note that it is most important for the argument that the π^- is captured from an atomic S state. The possibility of capture from P state has been calculated and found negligible.

§ 5. Charge Independence

We have now established that the spin of all the π mesons is zero, and that the charged mesons are pseudo-scalar. There are indications from the surprisingly large ratio of the two processes (*a*) and (*b*) and the smallness of (*e*) that π^0 is also pseudo-scalar (Marshak 1951), and we will take this as a working hypothesis from now on. If we further restrict ourselves (somewhat arbitrarily) to renormalizable interactions, the interaction energy density is

$$\begin{aligned} H=g_c(&\bar{\psi}_p\gamma_5\psi_n\phi_c+\bar{\psi}_n\gamma_5\psi_p\phi_c^*) \\ &+g_p\bar{\psi}_p\gamma_5\psi_p\phi_0+g_n\bar{\psi}_n\gamma_5\psi_n\phi_0. \end{aligned} \qquad \ldots\ldots(4.4)$$

ϕ_c is the charged meson field and g_c the corresponding coupling constant. We have arranged the meson and nucleon operators so that charge is conserved in each interaction (ϕ_c annihilates mesic charge, $\bar{\psi}_p$ creates charge in the nucleon field). The neutral meson field ϕ_0 is coupled independently to the neutron and proton fields with coupling constants g_n and g_p, which need not be equal.

This interaction still contains three arbitrary constants. In order to make it more specific we use the experimental result of the charge independence of nuclear potentials. It is known that if the spin and angular momentum states are the same, then apart from electromagnetic effects, the potentials between two protons, two neutrons, or a neutron and a proton, are all equal. If the potential is interpreted as the effect of a meson field, this clearly implies some relation between the couplings of the charged and neutral fields.

Let us consider this problem in lowest order perturbation theory. The graphs for nucleon–nucleon scattering are shown in figure 4.1. We have used double arrows to denote the motion of change, so that a double arrowed line denotes a proton, a single arrowed line, a neutron. For the p–p and n–n effect only the exchange of neutral mesons plays a part. For the n–p scattering there is an additional graph due to the charged meson, giving a charge-exchange term. Each

graph makes the same contribution to the scattering, and hence to the equivalent potential, apart from the coupling constants. Equating the coupling constants, as they appear in the three potentials, we obtain the conditions

$$g_p^2 = g_n^2 = g_c^2 + g_p g_n. \qquad \ldots\ldots(4.5)$$

These equations have two solutions:

$$g_p = g_n, \quad g_c = 0,$$
or
$$g_p = -g_n = g_c/\sqrt{2}. \qquad \ldots\ldots(4.6)$$

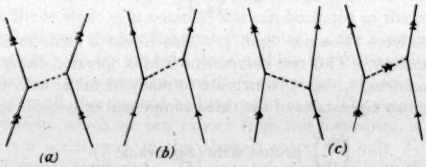


Figure 4.1. Nucleon–nucleon scattering. Double arrow denotes motion of charge. (*a*) p–p scattering. (*b*) n–n scattering. (*c*) n–p scattering, showing the ordinary and charge exchange terms.

The first solution implies that only the neutral meson field is coupled to the nucleon field, a result which is definitely excluded by experiment. We thus take the second solution. We now have an interaction depending on just one coupling constant,

$$H = g[\bar{\psi}_p \gamma_5 \psi_n \phi_c + \bar{\psi}_n \gamma_5 \psi_p \phi_c^* + \frac{1}{\sqrt{2}}(\bar{\psi}_p \gamma_5 \psi_p - \bar{\psi}_n \gamma_5 \psi_n)\phi_0]. \qquad \ldots\ldots(4.7)$$

This interaction gives charge independent nucleon potentials in the lowest order in an expansion in the coupling constant, which we know to be quite inadequate. It is not clear that it will give charge independence in general.

§ 6. Isotopic Spin

The complete solution to this problem was also given by Kemmer (1939). It involves the very formal concept of isotopic spin, but an important new fact comes out from this approach. As this proves very valuable in the analysis of high-energy meson experiments, it is well worth while to go into it in some detail.

In the non-relativistic theory of the electron of spin $\frac{1}{2}$, the wave function has two components, which may be written

$$\psi(x) = \begin{pmatrix} \chi_1(x) \\ \chi_2(x) \end{pmatrix}. \qquad \ldots\ldots(4.8)$$

This two-component wave function is operated on by the Pauli spin matrices

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \qquad \ldots\ldots(4.9)$$

The two eigenstates of the z-component of spin, $\frac{1}{2}\sigma_z$, are

$$\begin{pmatrix}\chi_1\\0\end{pmatrix}, \text{ spin 'up' (eigenvalue } \tfrac{1}{2})$$

and

$$\begin{pmatrix}0\\\chi_2\end{pmatrix}, \text{ spin 'down' (eigenvalue } -\tfrac{1}{2}).$$

In exact analogy with this, we combine the proton and neutron fields ψ_p and ψ_n into a single two-component field,

$$\psi = \begin{pmatrix}\psi_p\\\psi_n\end{pmatrix}, \qquad \ldots\ldots(4.10)$$

which is said to have 'isotopic' spin $\frac{1}{2}$. This is just a name to emphasize the mathematical analogy. This two-component field is operated on by the so-called isotopic spin matrices, τ_1, τ_2, τ_3, which are formally identical with the Pauli spin matrices. The two eigenstates of the three-component of isotopic spin $\frac{1}{2}\tau_3$ are

$$\begin{pmatrix}\psi_p\\0\end{pmatrix}, \text{ proton state (eigenvalue } \tfrac{1}{2}).$$

and

$$\begin{pmatrix}0\\\psi_n\end{pmatrix}, \text{ neutron state (eigenvalue } -\tfrac{1}{2}).$$

The actual charge is given by $\frac{1}{2}(\tau_3+1)$.

Now let us write the charged meson field in terms of its real and imaginary (Hermitian and anti-Hermitian) parts

$$\begin{aligned}\phi_c &= \frac{1}{\sqrt{2}}(\phi_1+i\phi_2)\\ \phi_c^* &= \frac{1}{\sqrt{2}}(\phi_1-i\phi_2)\\ \phi_0 &= \phi_3.\end{aligned} \qquad \ldots\ldots(4.11)$$

The factors $\sqrt{2}$ appear if the real fields ϕ_1 and ϕ_2 are normalized in the same way as ϕ_c. The final equation for ϕ_0 is just a convenient change of notation. In these new variables (dropping γ_5 for the moment, since it does not affect the argument and absorbing a factor $1/\sqrt{2}$ into the constant g) the interaction Hamiltonian deduced above may be written as a matrix product,

$$H = g(\bar{\psi}_p, \bar{\psi}_n)\begin{pmatrix}\phi_3 & \phi_1+i\phi_2\\ \phi_1-i\phi_2 & -\phi_3\end{pmatrix}\begin{pmatrix}\psi_p\\\psi_n\end{pmatrix}. \qquad \ldots\ldots(4.12)$$

By comparison with the Pauli matrices it may be seen that this is just

$$H = g\bar{\psi}\gamma_5(\tau_1\phi_1+\tau_2\phi_2+\tau_3\phi_3)\psi = g\bar{\psi}\gamma_5\boldsymbol{\tau}.\boldsymbol{\varphi}\psi. \qquad \ldots\ldots(4.13)$$

In the last equality we have combined the three meson fields into a single vector field $\boldsymbol{\varphi}$ with components (ϕ_1, ϕ_2, ϕ_3).

Let us now consider the implications of this rather elegant form for the interaction. It was explained in Part II that a vector field $\boldsymbol{\varphi}$, with components (ϕ_x, ϕ_y, ϕ_z),

describes an assembly of particles of spin 1. The components are deliberately labelled x, y, z, because they refer to the real space in which we live. The eigenstates of the z-component of the spin may be 1, 0 or -1. We have in our formalism a vector field $\boldsymbol{\varphi}$, with components (ϕ_1, ϕ_2, ϕ_3), with suffixes which refer to the charge and have nothing whatsoever to do with space directions x, y, z. However, just as in the nucleon case, we may say by analogy that our field $\boldsymbol{\varphi}$ has isotopic spin 1. Again, as in the nucleon case, and in analogy with ordinary spin, the eigenstates of the 3-component of the isotopic spin give the three charge states π^+, π^0 and π^-.

We now consider the total isotopic spin T of the interacting nucleon–meson system, and its 3-component T_3. In any interaction the total charge must be conserved. In the isotopic spin notation this can be stated as the conservation of T_3, in exact analogy with the familiar conservation of J_z, the z-component of total angular momentum. So far we have done nothing but express the conservation of charge, formally, as the conservation of the 3-component of ' spin ' in an abstract space. If this were all, there would be little point in it. The very important extra piece of information, which we can extract from this formalism, is that since the interaction H is a scalar, $\boldsymbol{\tau}.\boldsymbol{\varphi}$, in this abstract space, not only T_3 but the total isotopic spin T is also conserved. This is in exact analogy with the conservation of total angular momentum J, for the corresponding interaction in ordinary space $\boldsymbol{\sigma}.\mathbf{L}$.

Since the whole theory has been developed by analogy with the theory of spin, T, for any state, may be constructed in just the same way as the total spin J for an assembly of spin $\frac{1}{2}$ and spin 1 particles. For example one nucleon (i-spin $\frac{1}{2}$) and one meson (i-spin 1), combine to give a total isotopic spin of $\frac{3}{2}$ or $\frac{1}{2}$. The possible 3-components are $\pm\frac{3}{2}$, $\pm\frac{1}{2}$ for the first case, and $\pm\frac{1}{2}$ for the second case. The addition of $\frac{1}{2}$ to the 3-component gives the total charge of the state.

Similarly a state of two nucleons (the sum of two i-spins $\frac{1}{2}$) may have $T=1$ or 0. For the $T=1$ states the possible eigenvalues and eigenstates of T_3 are:

T_3	1	0	-1
eigenstate	pp	$\frac{1}{\sqrt{2}}$(pn+np)	nn

We can now return to the point from which this discussion started. The conservation of T is equivalent to the invariance of the theory under rotations in this abstract isotopic spin space. But clearly from the eigenstates given above, and our knowledge of the corresponding triplet spin states, we can transform the pp, nn and np states into each other by just such rotations. Therefore the interaction gives rise, quite generally, to a charge independent nucleon–nucleon potential. The result does not depend in any way on an expansion in powers of the coupling constant.

To summarize, we have found that, if we combine the proton and nucleon fields into a single two-component field ψ, and the three meson fields into a single vector field $\boldsymbol{\varphi}$, we may express the total charge formally in terms of the 3-component T_3 of a ' spin ' (isotopic spin) in an abstract space. This is always conserved for any interaction which we ever consider. For the particular interaction, equation (4.13), which gives charge independence, not only T_3, but also the total isotopic spin T is conserved. This gives us a new constant of the motion or, as we may say, a new ' good ' quantum number.

§ 7. The Meson–Nucleon Interaction

One final point is this. We have excluded the gradient type interaction on the grounds that it is not renormalizable. To be consistent we must make sure that the interaction we do decide on is renormalizable. To ensure this it is necessary to add a direct meson–meson interaction with arbitrary coupling constant (Matthews and Salam 1951). Thus the final interaction, which we have good reason to hope actually describes meson–nucleon effects, is†

$$H=g\bar{\psi}\gamma_5(\boldsymbol{\tau}.\boldsymbol{\varphi})\psi+\lambda\varphi^4. \qquad \ldots\ldots(4.14)$$

We now wish to explain why it has not yet been possible to show that equation (4.14) does give predictions in quantitative agreement with experiment. If g and λ were both small numbers ($g^2/4\pi\hbar c<\frac{1}{10}$, say), there would be no further difficulty. The methods of Feynman and Dyson described in Part III would be applicable, and it would be a simple matter to calculate the cross sections for various processes following from (4.14), and to check whether certain values of the coupling constants would give agreement with experiment. Unfortunately the constant g is not small, as anticipated in Part III, and this weak coupling approximation is not applicable.

If the π meson were scalar, and the interaction were of the form (see table 3.1)

$$H=g\bar{\psi}\psi\phi, \qquad \ldots\ldots(4.15)$$

it would still be possible to obtain approximate solutions even for large g. (We ignore the λ term and the isotopic spin for this discussion.) The reason is this. The basic steps in any interaction are illustrated in figure 3.3. These split into two distinct groups; firstly those for which the number of nucleons remains constant ((*a*) to (*d*)), and secondly those for which a nucleon–anti-nucleon pair is created or destroyed ((*e*) to (*h*)). For the scalar theory, (4.15), and for energies small compared with the nucleon rest mass, the pair creation (or annihilation) terms are small compared with those of the first kind and may be neglected. The result is that the number of nucleons in any process remains constant even in the virtual states. In meson–nucleon scattering for example, in this approximation, only one nucleon is involved even in the virtual parts of the interaction‡. The result is that the particle *assembly*, or operator, aspects of the nucleon field may be neglected. With this ' no pair ' approximation it is possible to calculate cross sections for strong ($g^2/4\pi\hbar c\gg 1$) or even intermediate ($g^2/4\pi\hbar c\simeq 1$) coupling strengths.

† If we abandon the condition that our interaction must be such that, in the limit of small g, it leads to reasonable (that is, finite) results after mass and charge renormalization, then all the restrictions of spin, parity, and charge independence are equally well met by

$$H=g\bar{\psi}\gamma_5\gamma_\mu\left(\boldsymbol{\tau}.\frac{\partial\boldsymbol{\varphi}}{\partial x_\mu}\right)\psi.$$

This is very definitely a second choice, since no one has yet produced a self-consistent way of deriving anything but sheer nonsense (infinite cross sections and infinite level shifts) from such an interaction, even in the limit of very weak coupling (small g). It is very doubtful whether such an interaction has any physical meaning. However, it cannot be dismissed out of hand, because the same could have been said of the electron–photon interaction in 1946 (before renormalization was understood), and this is now the basis of a theory in agreement with experiment to one part in a million!

‡ This is in marked contrast with the anomalous magnetic moment of the electron, for example, described in Part III, § 4, where the effects of virtual pairs are included.

Unfortunately, for the pseudo-scalar interaction it is the pair creation terms which predominate. Even meson–nucleon scattering for example becomes, in virtual states, a many-nucleon problem. The particle assembly, or operator, aspects of the field are essential. Thus for interaction (4.14) with the observed coupling strength neither of the two workable approximations ' weak coupling ' or ' no pair creation ' are applicable. No reliable quantitative calculations have yet been made of any of the physical consequences of such an interaction—not even on the assumption that λ is very small. If λ is also large, as it may well be, we are even further from a solution.

Part V. HIGH ENERGY EXPERIMENTS

By G. Feldman

In the previous parts we have seen how to develop a theory of interacting particles, in particular, the interaction between charged particles and photons, and the interaction between π mesons and nucleons. All of these theories require a knowledge of the mass, charge, intrinsic spin and intrinsic parities of the particles. Having established these quantities experimentally we can write down (as has been shown above) the possible theories which describe their behaviour. In this way, all experiments involving the electromagnetic interaction alone have been explained with the greatest accuracy. In the case of the meson–nucleon system we have been able to explain all of the qualitative results such as the possible reactions which may occur, the decay of the π^0 meson into γ-rays, etc. As a logical consequence to the previous part there should now be a report of the excellent quantitative agreement between theory and experiment in the interaction between mesons and nucleons. Unfortunately, this is not possible. We cannot solve the equations of meson theory in the same way as we do in electrodynamics. As has been pointed out, this is mainly due to the difference in size of coupling constants; $e^2/4\pi\hbar c = 1/137$, $g^2/4\pi\hbar c \simeq 10$; and electrodynamics is solved as a perturbation expansion in powers of $e^2/4\pi\hbar c$. Many attempts to solve the meson–nucleon equations are being made and perhaps will bear fruit in the near future.

Accordingly, we consider some of the experiments which have been performed in the past few years in order to see what regularities we can detect. A careful study of these experiments may be useful in giving the theorist some direction. First, we list the experiments which are most fundamental and are the ones the theorist would like to explain. It is also useful to picture these experiments in terms of the lowest order perturbation theory graphs, although there is no connection between the experimental results and those to which the graphs lead.

(i) Meson–nucleon scattering (figure 3.10(d))

$$\pi + \mathrm{N} \rightarrow \mathrm{N} + \pi,$$

where N is a nucleon and π a π meson.

(ii) Photo-meson production (figure 3.10(c))

$$\gamma + \mathrm{N} \rightarrow \mathrm{N} + \pi.$$

The energy required by the γ-ray in the laboratory system is 150 Mev at threshold.

(iii) Nucleon–nucleon scattering (figure 3.10(a))

$$\mathrm{N} + \mathrm{N} \rightarrow \mathrm{N} + \mathrm{N}.$$

(iv) Single and multiple meson production in nucleon–nucleon collisions (figures 3.10(b) and 5.1(a))

$$\mathrm{N} + \mathrm{N} \rightarrow \mathrm{N} + \mathrm{N} + s\pi,$$

where $s = 1, 2, 3, \ldots$ Thresholds are 290, 600, 900 Mev respectively.

(v) Meson production in meson–nucleon collisions (figure 5.1(*b*))

$$\pi+\mathrm{N}\rightarrow\mathrm{N}+s\pi$$

where $s=2, 3. \ldots$ Thresholds are 170, 365. . . . MEV respectively.

(vi) Multiple meson production in photo-nucleon collisions

$$\gamma+\mathrm{N}\rightarrow\mathrm{N}+s\pi$$

$s=2$, threshold 300 MEV (not observed as yet).

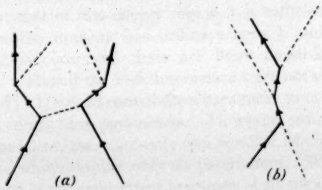


Figure 5.1. Meson production: (*a*) double production in N–N collision; (*b*) production in π–N collision.

As another preliminary it may be useful to mention the energies available and the types of machines producing them. The synchro-cyclotrons (at Chicago, Carnegie, Columbia etc.) produce nucleons up to about 450 MEV and π mesons up to about 250 MEV. The proton synchrotrons (mainly the one at Brookhaven) have reported experiments using nucleons up to 2300 MEV and π mesons up to 1500 MEV. The γ-rays (produced at Cornell, California Institute of Technology etc.) have produced results up to about 400 MEV.

The experiments which theorists have looked forward to with great interest are those on π–nucleon collisions. A good deal of data has now been amassed on the total and differential scattering cross sections for $\pi^{\pm}$ scattering on free protons. (For a complete set of references see de Hoffmann *et al.* (1954) and Bethe and de Hoffmann (1955).) The earliest and best experiments have been done with π energies up to 200 MEV and the reactions observed are the following

$$\left.\begin{array}{ll}\pi^{+}+\mathrm{p}\rightarrow\mathrm{p}+\pi^{+} & (a)\\ \pi^{-}+\mathrm{p}\rightarrow\mathrm{p}+\pi^{-} & (b)\\ \pi^{-}+\mathrm{p}\rightarrow\mathrm{n}+\pi^{0} & (c)\end{array}\right\} \quad \ldots\ldots(5.1)$$

Multiple meson production although possible (the threshold is 170 MEV) is very small and so these are effectively the only reactions observed in this energy range. (Any reaction involving production of γ-rays will likewise be unimportant since those cross sections will be down by a factor 1/137.) Reactions (5.1(*a*)) and (5.1(*b*)) are elastic scattering and (5.1(*c*)) is called charge exchange scattering. A measurement of the total π^{-} cross section will involve the sum of (5.1(*b*)) and (5.1(*c*)). Total π^{+}, π^{-} cross sections have been measured with mesons up to 1·5 KMEV and the results are shown in figure 5.2. Of course, at the higher energies multiple meson production takes place implying that other reactions in addition to the set (5.1) are involved.

The total cross-section curves exhibit some very interesting features. The first striking feature is the marked peak in the π^+ curve and a similar smaller peak in the π^- curve at about 200 MeV energy. There also seems to be a much broader peak in the π^- curve round 900 MeV. If we assume that the meson–nucleon interaction is charge independent the interpretation of these results fits into a consistent

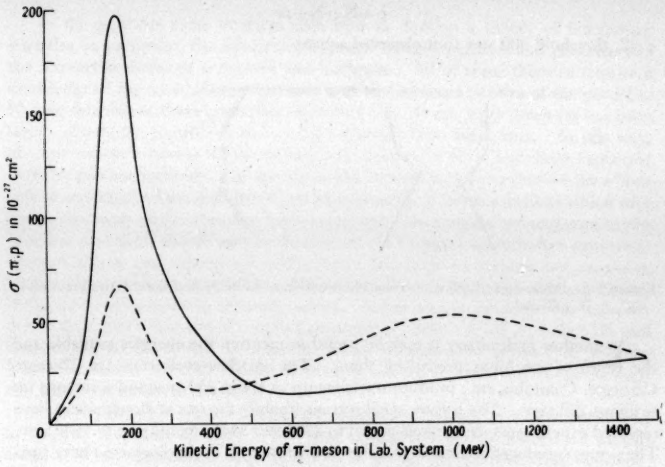


Figure 5.2. Experimental curves for π–p scattering.

picture. Let us, for the present, concentrate on the results up to about 200 MeV in which case the only reactions which enter are the set (5.1). Suppose that the π–nucleon interaction were not charge independent and in fact that there were no relation amongst the possible π–N experiments (except, of course, that arising from detailed balance), and that we perform all the possible experiments listed in (5.2).

$$\begin{aligned} &\pi^+ + \mathrm{p} \rightarrow \mathrm{p} + \pi^+,\ \pi^- + \mathrm{p} \rightarrow \mathrm{p} + \pi^-,\ \pi^0 + \mathrm{p} \rightarrow \mathrm{p} + \pi^0 \\ &\pi^+ + \mathrm{n} \rightarrow \mathrm{n} + \pi^+,\ \pi^- + \mathrm{p} \rightarrow \mathrm{n} + \pi^0,\ \pi^0 + \mathrm{n} \rightarrow \mathrm{n} + \pi^0 \\ &\pi^+ + \mathrm{n} \rightarrow \mathrm{p} + \pi^0,\ \pi^- + \mathrm{n} \rightarrow \mathrm{n} + \pi^- \end{aligned} \qquad \ldots\ldots(5.2)$$

There are eight possible experiments. Suppose, now, that they are done at very low energies, or in other words that only s angular momentum states enter into the picture. This means that each reaction would depend on one constant for its complete explanation. Thus, the set (5.2) would require eight constants for an interpretation of the results. However, if the interaction is charge independent, this implies certain relations amongst the reactions of (5.2) and greatly restricts the required number of parameters. By charge independence we mean that the π–nucleon reactions depend only on the total isotopic spin, that is, the results of experiments will not depend on the particular particles taking part nor on their charge but only on the *total* isotopic spin. We have seen in Part IV that a nucleon

has isotopic spin $\frac{1}{2}$ and the π meson has isotopic spin 1. Thus a π–N pair can exist in only *two* total isotopic spin states, $T=\frac{3}{2}$ and $\frac{1}{2}$. (Isotopic spin combines in the same way as ordinary spin.) Thus if the π–N interaction is truly charge independent, the set of experiments (5.2) can be explained not in terms of eight parameters but in terms of two. We were assuming that only s states of angular momentum were important. In general, of course, each angular momentum state will contribute two parameters (one for each of the two isotopic spin states).

We now take the differential cross section results of the set (5.1) which have been obtained up to about 220 MeV and analyse them into the different angular momentum states. It turns out that in this energy region it is sufficient to consider s and p states alone. Since the nucleon has ordinary spin $\frac{1}{2}$ and since only the *total* angular momentum is conserved there are three possible angular momentum states $s_{1/2}$, $p_{1/2}$, $p_{3/2}$. Thus if the π–N interaction were not charge independent the three experiments of (5.1) would contribute a parameter to each angular momentum state, which would imply nine parameters. However, assuming that the interaction is charge independent, there are only two possible isotopic spin states for each angular momentum, which implies only six parameters. We find, in fact, that we can explain the results of the experiment by means of six parameters only, indicating that charge independence of the π–N interaction is not in conflict with experiment. The parameters which we use to fit the result could be the scattering amplitudes in the different states, but we find it more convenient to use the parameters called the phase shifts. These are completely analogous to those used in nuclear physics. The existence of a peak in the π–p cross sections at about 200 MeV indicates that this may be a resonance (implying an excited state of a nucleon). The peak is truly a resonance if some phase shift passes through 90°. When we analyse the experimental results into phase shifts we find that there are, in fact, two sets which can explain the results; these are usually termed the Fermi and Yang sets. Neither set can be eliminated by making angular distribution experiments alone. However, if the polarization of the recoil proton is measured, we could probably eliminate one set. At present, theorists are inclined to believe the Fermi set and this gives a resonance in the $T=\frac{3}{2}$, $p_{3/2}$ states. Such a resonance in the $T=\frac{3}{2}$, $p_{3/2}$ state is predicted by the so-called strong coupling theory (but this theory is not very reliable from many other points of view). The existence of such a resonance is also in agreement with other results. Since the third component of isotopic spin for the proton is $+\frac{1}{2}$ and for the π^+ is $+1$, the total third component of the (π^+, p) system is $T_3=\frac{3}{2}$. This can only come from a total T of $\frac{3}{2}$. Therefore the (π^+, p) system is a pure $T=\frac{3}{2}$ state. However, the (π^-, p) pair has $T_3=-\frac{1}{2}$ and so comes from a mixture of $T=\frac{3}{2}$ and $\frac{1}{2}$. In fact the (π^-, p) system is $\frac{1}{3}$ in the $\frac{3}{2}$ state. Thus we would expect the π^+ curve to have a greater value at resonance than the π^- curve, and, provided scattering in the $T=\frac{1}{2}$ state is small, to be just three times the π^- curve. This result is, in fact, observed. From resonance theory in nuclear physics we know that if all states are swamped out by the resonance state, the maximum total cross section is given by

$$\sigma_{\max}=\frac{2J+1}{(2j+1)\,(2s+1)}\,4\pi\left(\frac{\lambda}{2\pi}\right)^2 \qquad \ldots\ldots(5.3)$$

where J is the spin of the compound nucleus and j, s are the spins of the target and bombarding particles respectively and $\lambda/2\pi$ is the wavelength in the rest system. In this case $J=\frac{3}{2}$, $j=\frac{1}{2}$, $s=0$.

Therefore we should have

$$\left.\begin{aligned}\sigma_{\max}(\pi^+, \mathrm{p}) &= 8\pi\left(\frac{\lambda}{2\pi}\right)^2\\ \sigma_{\max}(\pi^-, \mathrm{p}) &= \frac{8}{3}\pi\left(\frac{\lambda}{2\pi}\right)^2\end{aligned}\right\} \quad \ldots\ldots(5.4)$$

This is in agreement with the experimental results.

We can sum up by saying that there exists an excited nucleon state, or isobar, with energy about 300 MeV above the ground state of the nucleon, with a lifetime about 10^{-23} second, with ordinary spin $\frac{1}{2}$ and which can exist in four possible charge states, doubly positive, positive, neutral and negative. (The excitation energy is 300 MeV because the nucleon absorbs the total energy of the π meson to form the isobar; the mass of the π meson is 140 MeV and a 200 MeV meson in the centre-of-mass system has 160 MeV kinetic energy. The lifetime of the isobar is 10^{-23} second because the width of the resonance is of the order of 100 MeV and $\tau = \hbar/\Gamma$ where Γ is the width and τ the lifetime. There are four charge states because its isotopic spin is $\frac{3}{2}$ and the charge is given by $T_3 + \frac{1}{2}$.)

If we look again at figure (5.2) there appears what might be another very broad resonance in the neighbourhood of 900 MeV. This second peak appears in the π^- curve but not in the π^+ curve which would imply that the resonance is in a $T = \frac{1}{2}$ state. However, no angular distribution experiments have as yet been made and analysed in this energy region and they will be extremely complicated due to the occurrence of inelastic processes (multiple meson production).

Now let us turn to other processes to see how this nucleon isobar (of 300 MeV excitation) manifests itself. As is pointed out above we think of π–N scattering as proceeding through the isobaric state (at least in the neighbourhood of 200 MeV π mesons). That is we think of the reaction as

$$\pi + \mathrm{N} \rightarrow \mathrm{N}^* \rightarrow \mathrm{N}' + \pi', \quad \ldots\ldots(5.5)$$

where N^* is the isobar (or ' compound ' nucleon). Accordingly, if we perform an experiment on photo-meson production we should expect the following to take place:

$$\gamma + \mathrm{N} \rightarrow \mathrm{N}^* \rightarrow \mathrm{N}' + \pi. \quad \ldots\ldots(5.6)$$

The fact that in the intermediate state we have formed N^* again implies the occurence of a resonance in photo-meson production, at about 300 MeV γ-ray energy. The experiment on π^+ production is shown in figure 5.3 which indeed displays the correct features (the results are quoted by Bethe and de Hoffmann 1955).

The next most interesting process is nucleon–nucleon total cross sections. The total cross section consists of the elastic part and the meson production (single or multiple) parts. It is conceivable that we might expect single meson production to take place largely as below

$$\mathrm{N} + \mathrm{N} \rightarrow \mathrm{N}' + \mathrm{N}^* \rightarrow \mathrm{N}' + \mathrm{N}' + \pi. \quad \ldots\ldots(5.7)$$

In this case the intermediate state consists of the isobar plus another nucleon and so the cross-section curve will not simply show a peak. The fact that there are two particles in the intermediate state means that any amount of kinetic energy can be shared. Thus, once there is sufficient energy to produce N^*, any energy greater

than this amount will also produce N*. We would then expect the cross section to rise sharply at a certain value and to remain up as shown in figure 5.4. From kinematical considerations we find that this rise should occur when the incoming nucleon has about 650 MeV.

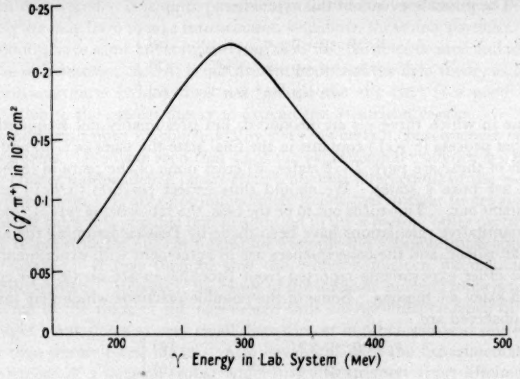


Figure 5.3. Experimental curve for photo-production of π^+ mesons.

The total cross section in p–p collisions has been performed at Brookhaven (Shapiro *et al.* 1954) up to energies of about 1200 MeV and the results (figure 5.4) are what we might expect.

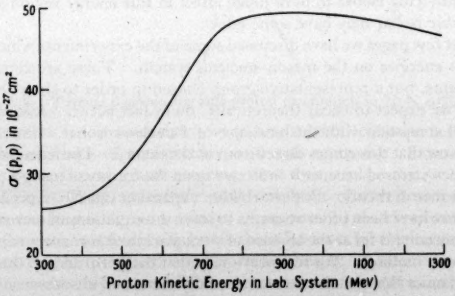


Figure 5.4. Experimental curve for total p–p scattering.

When we use nucleons of even greater energy (1300 MeV and more) we might also expect the following reaction to be very important:

$$N+N\rightarrow N^*\rightarrow N^*\rightarrow N'+N'+\pi+\pi, \qquad \ldots\ldots(5.8)$$

where we now have two isobars in the intermediate states. Other experiments performed at Brookhaven (Fowler *et al.* 1954) on neutron–proton collisions confirm this picture very well. The experiment consists of allowing a beam of neutrons (energy from 1 to 2·3 kMeV) to enter a hydrogen diffusion chamber and then to look for three-pronged events, events in which three charged particles emerge from a point. The possible events of this type are

$$\left.\begin{aligned} n+p &\rightarrow p+p+\pi^- \quad (a)\\ &\rightarrow p+p+\pi^-+\pi^0 \quad (b)\\ &\rightarrow n+p+\pi^++\pi^- \quad (c)\end{aligned}\right\} \qquad \ldots\ldots(5.9)$$

and some in which three π's are produced, but presumably not frequently. We notice that process (5.9(c)) contains in the final state the pairs (π^+, p) and (π^-, n) and each of these are pure $T=\frac{3}{2}$ states, whereas none of the pairs in (5.9(a)) and (5.9(b)) are pure $\frac{3}{2}$ states. We should thus expect reaction (5.9(c)) to be the predominant one. This turns out to be the case, the ratios being (c):(b):(a)=3:1:1. Some quantitative calculations have been made by Peaslee assuming the existence of such an isobar, and the consequences are in agreement with experiment.

Some other experiments reported from Brookhaven are on (π^-, p) collisions using 1·5 kMeV π^- mesons. Some of the possible reactions which may take place and are observed are

$$\left.\begin{aligned} \pi^-+p &\rightarrow p+\pi^- \quad (a)\\ &\rightarrow p+\pi^-+\pi^0 \quad (b)\\ &\rightarrow n+\pi^-+\pi^+ \quad (c)\\ &\rightarrow n+\pi^0 \quad (d)\\ &\rightarrow n+2\pi^0 \quad (e)\\ &\rightarrow p+\pi^++2\pi^- \quad (f)\end{aligned}\right\} \qquad \ldots\ldots(5.10)$$

However, as we see from the total cross-section results in figure (5.2) we should not expect the low-lying isobar to have much effect in this energy region although the second possible isobar may have some effect.

In the last few pages we have discussed some of the experiments which have been done at high energies on the meson–nucleon system. These are certainly not all the experiments, but a representative group chosen in order to show that the reactions which we expect to occur theoretically, do in fact occur. Also these experiments are all consistent with the existence of a nucleon isobar. It is now for the theorist to show that this comes directly out of the theory. The existence of such an isobar has been stressed because it indicates again the futility of using a perturbation expansion in meson theory. A perturbation expansion can never produce such an isobar. There have been other attempts to solve the equations of meson theory and the most important (as far as the amount of work put into it is concerned) is called the Tamm–Dancoff method. We saw previously that the perturbation theory implied in electrodynamics that the electron occasionally emits and absorbs a single photon. In meson theory, however, due to the strength of the coupling, the nucleon is surrounded by many virtual mesons (a ' cloud ' of mesons). The Tamm–Dancoff method takes this effect into account approximately by assuming that there are not more than a certain number of virtual mesons or pairs present at one time. The usual basis for the approximation is mathematical simplicity rather than physical reality and it is true to say that the Tamm–Dancoff method has not yielded many

satisfactory results. Nothing will be said about more recent attempts to solve the equations of meson theory because no quantitative results have as yet been obtained.†

Up to this point we have been discussing the π-meson–nucleon interaction alone and the results of experiments using these particles. While we are still trying to explain these π–N experimental results, new particles called the Λ's, τ's, θ's etc. have been discovered. It is quite possible that in order to explain the π–N results adequately we may have to take into account, somehow, these new particles. They will, no doubt, have some effect by entering as virtual particles created and absorbed during the π–N interaction. It is our present hope that the field theory as outlined in the previous parts (which does not include the Λ's etc.) is a good enough approximation to the correct theory to explain the π–nucleon results.

This discussion must be concluded by stressing certain consequences of theory and experiment. We have seen that all of the experimental evidence is consistent with the charge independence (apart from Coulomb interaction) of the π-meson–nucleon interaction. It does not immediately follow that nuclear forces are charge independent, since it is possible that in addition to the π meson, other particles may be partly responsible for nuclear forces, for example the τ meson (mass about 1000 electron masses). It may then be that the τ-meson–nucleon interaction is not charge independent. However, the range of the nucleon–nucleon force is inversely proportional to the mass of the meson and thus charge independence of nuclear forces might break down at very small distances, or in other words at much higher energies than occur inside nuclei. Also consistent with the experimental results is the existence of a nucleon isobar which has an extremely short lifetime, about 10^{-23} second, but which produces resonance effects in scattering curves.

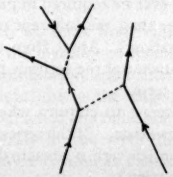


Figure 5.5. Typical diagram for anti-proton production in N–N collision.

Finally, it must be stressed that a very fundamental consequence of present theories is the existence of anti-protons (and anti-neutrons). Figure 5.5 is the lowest order perturbation graph showing how they are produced in nucleon–nucleon collisions. They apparently have not been observed as yet, but this should not be too surprising for a number of reasons: (i) the threshold energy for production in nucleon–nucleon collisions is about 5·6 kMev for the incoming nucleon (although they can be produced slightly more cheaply in other ways) and this energy, until very recently, was available only in cosmic rays; (ii) once having been produced they would look just like protons and would be most easily identified on coming to rest and being annihilated. All of these considerations make their possible observation up to this time rare. However, one event observed by Rossi and his collaborators (Bridge *et al.* 1954) may conceivably be an example of an anti-proton.

† Since these lectures were given some quantitative progress has been made in understanding the low energy π–N data.

Part VI. FINAL REMARKS. THE NEW PARTICLES

By R. E. Peierls

§ 1. Present State of Electromagnetic Theory

Up to now we have reviewed the principles of the quantum theory of fields, and their application to the interaction of charged particles with electromagnetic radiation and to the interaction between nucleons and mesons.

On the first subject, the position is fairly satisfactory, in the sense that there exists a unique answer to any problem involving only electromagnetic interactions, and that so far this answer has been confirmed by experiment to an impressive accuracy. It is a reasonable, though not proved, conjecture that on improving the accuracy both of experiment and calculation there would still be agreement, except that in any problem there are bound to be corrections due to processes involving the virtual production of mesons and nucleon pairs, and these would require a full understanding of meson field problems.

One reservation should here be made. The theoretical answers always appear in the form of a series expansion in powers of the fine structure constant, which is 1/137. It has been stressed that the success of the theory is largely due to the small value of this constant. Nevertheless it has not been proved that the series is convergent, and there are, in fact, indications that it is not. If it is not a convergent series, the good agreement so far obtained would suggest that we are dealing with a semi-convergent series, and that it can be used to give answers to a much higher degree of accuracy than would ever be required in practice. (In fact, the labour in computing higher terms, rather than the inherent errors of the asymptotic series, is likely to form the real limitation.) Apart from the meson corrections already mentioned, the asymptotic character of the series is probably unimportant until we reach something like its 137th term.

However, it is of some interest to enquire whether an exact solution to the present quantum field equations exists. If the series is semi-convergent it does not in itself define an exact solution (though it presumably approaches very closely to the exact solution, if the latter exists).

This is the place at which the occurrence of divergent integrals, and hence of infinite renormalization constants, still causes conceptual difficulties. If there were no infinities there would be no difficulty in deciding whether the equations with which we are faced had a reasonable set of solutions. Because of the infinities we must interpret the mathematical process of solving the equations in a manner in which every infinite result is removed by an adjustment of the physical constants. How to do this has to be specified in relation to the actual procedure of solving the equations, and the only procedure fully investigated is the series expansion.

That these doubts are not entirely academic can be illustrated by considering some simplified problems in which exact solutions are obtainable. Some of these models show new, and physically unreasonable, features which occur in the exact solution, but are lost in the series expansion. For example, one may consider a classical field theory in which the quantum constant $\hbar$ is neglected. If here the mass renormalization is carried out, new unreasonable results are obtained if the 'electron radius' is so small that the self-mass δm exceeds the mass of a real particle,

so that the ' bare ' mass M_0 has to be negative. Since the series expansion treats the mass correction as if it were small, it does not show up these difficulties. It should be stressed that such a classical model has no similarity to the real physical situation, since the fine structure constant contains $\hbar$ in the denominator, so that $\hbar \to 0$ takes us to the case of infinitely large fine structure constant as compared with the actual small value. The model should serve only to show up the mathematical possibilities.

It was pointed out in Part I of this series that in quantum theory one is usually dealing only with logarithmic divergences, so that the renormalization constants do not become appreciable unless the integrations are carried to extremely small wavelengths. If it should turn out that the point interactions of the present theory through the divergent renormalization constants would lead to physically unreasonable results, we would have to expect that new factors (such as the meson terms already mentioned) would modify the situation for very small distances or very short wavelengths. There is no reason to expect that this change would result in any noticeable changes in the practical results so far obtained, but we should then have to regard the present theory as an incomplete scheme rather than a complete and rigorous theoretical framework.

§ 2. Note on the Classical Limit for Waves and Particles

Before leaving the case of electrons and electromagnetic fields I would like to interpose a remark concerning a simple question which is raised frequently.†

In the previous discussion we have stressed the equivalence and inter-relation between the wave and particle aspects both of electrons and photons. Is it purely an accident that in the development of physics electrons first appeared as particles, and light as waves?

In answering this question we must first remember that there is a range of phenomena, those covered by the methods of geometrical optics, in which both for light and for electrons, the simultaneous application of wave and particle concepts is possible. In the case of light, this is the stage of the early ' emission theory ' which is adequate to describe refraction and reflection, but not diffraction phenomena.

This was followed historically by the wave theory, culminating in Maxwell's equations, in which the fields are treated as classical variables. These equations are particularly appropriate for dealing with the propagation of radio waves, in which not only the intensity but also the phase of the wave is conveniently measurable with existing instruments. In such measurements the uncertainty principle presents no practical limitation.

Now according to the uncertainty principle, exact measurements of the intensity and the phase are mutually exclusive. With some oversimplification we may write

$$\Delta N \Delta \gamma \gtrsim 1. \qquad \ldots\ldots (6.1)$$

Here N is the number of photons in one mode of the radiations field, γ the phase. Hence no information about the phase is possible, unless, after the measurement, there is some uncertainty in the number of photons, i.e. unless the measuring apparatus is capable of creating or annihilating photons. This is true of any field

† This question was asked during the discussions at the Physical Society Autumn meeting; the points raised by many other such questions have been incorporated in the text of these articles and the authors wish to acknowledge the help which these constructive questions have given in the presentation of the material.

measuring device, indeed as was shown in Part III, the basic interaction between electromagnetic waves and matter consists of creation and annihilation processes.

But to apply classical equations, we should be in the possession of reasonably accurate knowledge of the fields, i.e. the error in the phase should be small compared with 2π, and the error in N small compared with N itself. This requires, according to (6.1), that N should be a large number. To be quantitative, N must here be interpreted as the density of photons in the cube of a wavelength. One easily verifies that in the field of a real radio transmitter N is indeed very large, so that this case meets all the requirements of a classical field situation.

On the other hand, the same does not hold for electrons. In the first place they are charged, so that no experiment can create or destroy a single electron. Hence the phase of the electron wave function is, by (6.1), always unobservable. (We can create or destroy electron–positron pairs, but this makes possible only the knowledge of the phase of the product of the two wave functions representing the two kinds of particles.) Secondly, we are dealing with particles of a finite rest mass, so that the creation of one of them, even for a neutral particle, requires at least the finite energy mc^2. An instrument which is capable of doing this, must therefore have a very drastic interaction with the field. This, in itself, does not make such a measurement impossible, but it makes it more difficult. Thirdly we can never meet, in the case of electrons, the condition that N be large, since by Fermi–Dirac statistics the number of particles in one mode of motion N can at most be 1.

Whereas, therefore, we can never realize conditions in which a classical field would be adequate for describing electrons, we do meet conditions where classical particle dynamics is appropriate for them.

This is true firstly because their finite rest energy means that in low-energy interactions no pair creation takes place, so that the number of electrons does not change. However complicated the field of force, we can always count the electrons and get a constant answer. Indeed, the experiments associated with the discovery of the electron involved mainly counting electrons through the effect of their discrete electric charge.

In most experiments on the mechanics of electrons (e.g. deflection in a cathode-ray tube) we are in conditions of geometrical optics, when the use of wave packets would serve as well as classical mechanics. However, we may use the concept of the position of an electron even beyond the limits of geometrical optics. For instance, we may look for the electron within an atom, and the results of such observations will be distributed according to the statistical prediction of wave mechanics. (In each such observation the state of the atom must of course be drastically changed, according to the uncertainty principle.) This is done, in fact, in determining the electric density inside a crystal by x-ray diffraction. Thus in wave mechanics, the electron density remains a useful concept. It was stressed in Part II that no exact expression for the particle density can be given for particles obeying Bose–Einstein statistics, such as π mesons or photons. In the case of π mesons we may define a charge density, but a given charge can still be made up in various ways by positive and negative particles, so that there is no simple relation between charge density and particle density.

Both for photons and mesons we can work with the energy density. For mesons we may be dealing with a non-relativistic situation in which the kinetic energy is negligible compared with their rest energy. (This applies for example

to a mesic atom of not too high atomic number.) In this limit the number of particles is just the energy, divided by the rest energy of one particle. Then the particle density becomes a useful concept, and one-particle dynamics is applicable to the non-relativistic motion of mesons.

In the case of electrons, the one-particle view is really also limited to the non-relativistic case, since otherwise the transitions to states of negative energy, which have to be interpreted as related to pair production, as we saw in Parts II and III, play a part. Hence one-particle dynamics remains a useful picture for the non-relativistic motion of electrons and mesons. It has no application to photons, which have no rest mass and always move with the velocity of light.

§ 3. Difficulties of Meson Theory

Returning now to our main theme, we recall that the understanding of the meson–nucleon interaction has not progressed nearly as far as that of electron–photon problems, and that the principal difference lies in the greater strength of the coupling.

As a result, the expansion in powers of the coupling constant, which was so powerful in the electromagnetic case, is useless in this problem, and there exists no reliable mathematical procedure for investigating the solution. The doubts relating to the part played by infinite renormalization constants are here also far from academic. If the series in powers of the coupling constant converges, it is useless for computing purposes; if it is semi-convergent (which appears likely) it does not even define the answer to any reasonable accuracy.

It is not even clear whether the numerical value of the coupling constant can be given a clear and unambiguous meaning. It was stressed in Part IV that the nucleon–meson coupling constant g is, like the electric charge e, affected by renormalization. The charge e can easily be defined in physical terms, both because it is conserved, and because it can be related to experiments describable in purely classical terms. Neither of these statements apply to g, and g can be related to experiments only by using the quantum theory of fields itself. Several ways have been suggested of determining g directly from experiments with low energy π mesons, using as little detail of the theory as possible, but it has not yet been proved that different methods lead to comparable answers on the basis of renormalized theory.†

As Dr. Matthews explained in Part IV, the theory contains, besides the coupling constant g, another constant λ, which describes the (renormalized) interaction between π mesons. In this case we do not even have any clear experimental evidence which would suggest a way of determining this constant. There is no doubt that it must be determined, directly or indirectly, before a satisfactory application of the theory is possible.

As far as our present knowledge goes, the particles so far mentioned, i.e. photons, electrons, nucleons and π mesons with the corresponding anti-particles, might represent the whole of physics, and we have discovered no reason (though reasons may well exist) why the description of their interaction should not be possible without bringing in any further particles or interactions. To be sure we have

† Since the meeting at which these lectures were given, the author has attended the Fifth Annual Conference on High-Energy Physics in the University of Rochester at which new work on this problem was reported which seemed to encourage a slightly more optimistic outlook.

known for long of the phenomenon of beta decay, and of the decay of the π meson, which demonstrates the existence of new types of interactions, and of neutrinos and μ mesons. But these interactions are so weak (i.e. the corresponding coupling constants are much less even than the fine structure constant) that it would be tempting to try a consistent description of nature in which these were neglected.

§ 4. New Particles of Surprisingly Long Life

However, this situation has been changed by the discovery, within the last few years, of a bewildering variety of new particles. The situation in this field resembles the state of our knowledge of the elements before Rutherford, when we had to accept a catalogue of 92 or so elements with all their properties as given by nature without being able to derive them from any more basic principle. In that case the discovery of the structure of the atom and of the quantum laws of motion of the atomic electrons provided the key. It is to be expected that the new particles also will be seen as logical consequences of a more general set of physical laws. It is not likely that this will take the form of regarding these particles as compound structures, built mechanically of simpler constituents. The reason for this belief is that any attempt to describe such systems involves interaction energies comparable with the rest energies of the constituents. In such circumstances we cannot expect the number of elementary particles to have a simple meaning and we would be forced, as in all high-energy problems, into a description in terms of quantized fields.

It is possible that some of the new particles can be explained as manifestations of a smaller number of fields and their interactions, but the degree of variety in the properties and interactions of these particles is such that a very small number of basic fields is unlikely to suffice.

This is not the occasion to give a list of all the known particles (see Rochester and Butler 1953 and Rossi 1955 for recent reports on this field), but for our purpose it will be sufficient to mention two. One of these is the Λ^0 particle, which decays according to the scheme

$$\Lambda^0 \rightarrow \mathrm{p} + \pi^- \qquad \ldots\ldots (6.2)$$

where p stands for a proton and π^- for a negative π meson. In this reaction an energy of about 37 Mev is released. The other is the θ^0 particle, decaying as

$$\theta^0 \rightarrow \pi^+ + \pi^- \qquad \ldots\ldots (6.3)$$

with an energy release of about 214 Mev. Both these particles are produced quite profusely in high-energy collisions and the production cross sections are not small compared with the quantity which one would expect to occur dimensionally, namely $(\hbar/m_\pi c)^2$.

On the other hand, both particles are relatively stable, their mean lives being of the order of 10^{-10} second. This should be compared with the natural time scale which would occur in such a theory, which is determined by the time $\hbar/m_\pi c^2$, which is of the order of 3×10^{-24} second.

This argument indicates that the production of these particles is the result of a very intense interaction, whereas their decay by the reactions (6.2), (6.3) is caused by an extremely weak coupling. An immediate consequence of this is that the reaction by which these particles are produced cannot be the direct inverse of (6.2) or (6.3). It is in fact not easy to postulate any other process for the production

which would not, as an indirect result, also accelerate the decay beyond the observed rate.

This is reminiscent of the earlier situation with the μ mesons in cosmic radiation. Their abundance showed that they were produced as a result of a strong interaction, yet they seemed to interact very weakly with nuclei. In this case the discrepancy was resolved by the discovery of the π meson. The π meson is strongly coupled to nucleons, and therefore easily produced. It decays slowly, the μ meson being one of the decay products.

A similar hypothesis of a ' heavy parent ' is not likely to resolve the difficulty as regards the new particles. Suppose, for example that the Λ particle was the decay product of some intermediary X produced by the reaction

$$\mathrm{N}+\pi\rightarrow\mathrm{X} \qquad \ldots\ldots(6.4)$$

where N stands for a nucleon, π for a neutral or charged π meson. To give abundant production (6.4), must be the result of a strong coupling. X then decays

$$\mathrm{X}\rightarrow\Lambda+\mathrm{Y} \qquad \ldots\ldots(6.5)$$

where Y may be a π meson or any other light particle. The difficulty is now this: There are two competing reactions for the decay of X. Besides (6.5) it may decay by the inverse of (6.4). It has enough energy for this, since it must be heavier than the Λ (otherwise (6.5) could not take place) and we know from (6.2) that this is more than the mass of the left-hand side of (6.4). Hence if (6.5) were a weak process, the X would predominantly decay by the inverse of (6.4), and no Λ would be formed. If (6.5) is strong, then Λ could go into a virtual state

$$\Lambda\rightarrow\mathrm{X}+\overline{\mathrm{Y}} \qquad \ldots\ldots(6.6)$$

$\overline{\mathrm{Y}}$ being the anti-particle to Y, and hence indirectly to the left-hand side of (6.4). This would still give too short a life for Λ. One can consider more complicated variants of this sequence, but none are capable of removing the discrepancy of a factor of 10^{-13} or so in the mean life.

The most convincing explanation which has been suggested (cf. Pais 1952) is that Λ^0 and θ^0 are not produced singly, but in pairs, for example

$$\mathrm{N}+\pi\rightarrow\Lambda^0+\theta^0 \qquad \ldots\ldots(6.7)$$

or

$$\mathrm{N}+\mathrm{N}\rightarrow\Lambda^0+\Lambda^0. \qquad \ldots\ldots(6.8)$$

The inverse of these strong reactions would relate to collisions between two new particles, and therefore would not be of practical interest.

It is interesting that in experiments at Brookhaven examples of the simultaneous production of such particles in the same event have been seen, but the evidence is by no means conclusive to rule out the occurrence of single production.

If single production were found to take place abundantly, the theory would be in great difficulty. If it does not occur this fact has to be listed as another qualitative success of simple theoretical reasoning.

§ 5. Different Types of Coupling

Accepting for the moment the latter position, we find therefore that we are confronted with some strong couplings including in particular those responsible for

the production processes such as (6.7) and (6.8) as well as the π–nucleon interaction, and some very weak interactions including the decay processes as (6.2) and (6.3) as well as the couplings responsible for beta decay and for the decay of the π and μ mesons. (Perhaps gravitational effects should also be listed under this heading.) The couplings of the first type must have interaction constants of the order of unity, those in the second list must be at least 10^{10} times smaller. Electromagnetic effects, with the coupling constant 1/137, occupy an intermediate position, but are closer to the ' strong ' interactions.

Finally, there is a class of processes which would be compatible with all known conservation laws, but which are known not to take place at all. An example would be

$$N+N\rightarrow 2\pi \qquad \ldots\ldots(6.9)$$

(e.g. two protons turning into two positive mesons).
This reaction would be violent exothermic. If it could take place even very slowly, very general thermodynamic arguments would show that all nuclei would disintegrate into π mesons, and this would go entirely in one direction. The fact that matter is stable is therefore an *immediate* proof that (6.9) is absolutely prohibited.†

The same applies for reactions in which two Λ particles or a Λ and a nucleon are destroyed, since, via the virtual transformation of a nucleon into a Λ, this could also lead to the decay of normally stable nuclei. We are therefore led to postulate a firm conservation of the number of nucleons, where for this purpose Λ particles and others occurring in reactions like (6.2), (6.7) or (6.8) have to be counted as nucleons, and anti-particles count negatively.

A future theory of the various types of particles will have to explain the rules by which the different possible couplings are grouped into the ' strong ', ' electromagnetic ', ' weak ' and ' absolutely forbidden ' types.

A very interesting attempt to formalize these selection rules has recently been made by Pais and Gell-Mann (1955) who link this with the known rule of charge-independence (Part IV) in which a formal analogy with rotations in an unspecified space is set up. By extending this analogy, these authors succeed in reproducing some of the features of the list of particles and interactions as so far disclosed by observation. The inclusion of the details would exceed the scope of this brief survey.

I have mentioned the new particles and some of the problems relating to them mainly as an indication of the magnitude of the unsolved problems. If one remembers the difficulties with which we were confronted in trying to handle the one strong coupling between two fields, the prospect of having to include a larger number of fields and several strong interactions looks, for the present, discouraging to the theoretician.

However, in this respect we are only at the beginning of the story, and it would not be for the first time that a new brilliant idea would bring not only new insight, but a new simplicity into an apparently obscure and complicated situation.

† The word ' absolute ' is of course an exaggeration. However if one estimates the reaction rate which would be compatible with the age of the universe, one finds the mean life for this reaction extremely small even by comparison with the ' weak ' couplings mentioned before.

References

ANDERSON, C. D., and NEDDERMEYER, S. H., 1937, *Phys. Rev.*, **51,** 884.

BETHE, H. A., and DE HOFFMANN, F., 1955, *Mesons and Fields* (Evanston, Ill.: Row, Peterson Co.).

BRIDGE, H. S., COURANT, H., DE STABLER, H., and ROSSI, B., 1954, *Phys. Rev.*, **95,** 1101.

CARTWRIGHT, W. F., RICHMAN, C., WHITEHEAD, W., and WILCOX, H., 1951, *Phys. Rev.*, **81,** 652.

CLARK, D. L., ROBERTS, A., and WILSON, R., 1951, *Phys. Rev.*, **83,** 649.

DIRAC, P. A. M., 1927, *Proc. Roy. Soc.* A, **114,** 243, 710; 1928, *Ibid.*, **117,** 610.

DURBIN, R., LOAR, H., and STEINBERGER, J., 1951, *Phys. Rev.*, **83,** 646.

DYSON, F. J., 1949, *Phys. Rev.*, **75,** 486, 1736.

FERMI, E., 1932, *Rev. Mod. Phys.*, **4,** 131.

FEYNMAN, R. P., 1948, *Rev. Mod., Phys.*, **20,** 367; 1949, *Phys. Rev.*, **76,** 749, 769.

FOWLER, W. B., SHUTT, R. P., THORNDIKE, A. M., and WHITTEMORE, W. L., 1954, *Phys. Rev.*, **95,** 1026.

HEISENBERG, W., and PAULI, W., 1929, *Z. Phys.*, **56,** 1.

DE HOFFMANN, F., METROPOLIS, N., ALEI, E. F., and BETHE, H. A., 1954, *Phys. Rev.*, **95,** 1586.

KARPLUS, R., and KROLL, N. M., 1950, *Phys. Rev.*, **77,** 536.

KEMMER, N., 1938 a, *Proc. Roy. Soc.* A, **166,** 127; 1938 b, *Proc. Camb. Phil. Soc.*, **34,** 354.

KOENIG, S., PRODELL, A. G., and KUSCH, P., 1951, *Phys. Rev.*, **83,** 687.

LAMB, W. E., 1951, *Rep. Progr. Phys.* (London: Physical Society), **14,** 19.

LAMB, W. E., and RETHERFORD, R. C., 1950, *Phys. Rev.*, **79,** 549.

LATTES, C. M. G., OCCHIALINI, G. P. S., and POWELL, C. F., 1947, *Nature, Lond.*, **160,** 453.

MARSHAK, R. E., 1951, *Rev. Mod. Phys.*, **23,** 137.

MATTHEWS, P. T., and SALAM, A., 1951, *Rev. Mod. Phys.*, **23,** 311.

PAIS, A., 1952, *Phys. Rev.*, **86,** 663.

PAIS, A., and GELL-MANN, M., 1955, *Proceedings of the* 1954 *Glasgow Conference on Nuclear and Meson Physics* (London: Pergamon Press).

PANOFSKY, W. K. H., AAMODT, L., HADLEY, J., and PHILLIPS, R., 1950, *Phys. Rev.*, **80,** 94.

PAULI, W., 1941, *Rev. Mod. Phys.*, **13,** 213.

PAULI, W., and WEISSKOPF, V. F., 1934, *Helv. Phys. Acta*, **7,** 709.

PEASLEE, D. C., 1954, *Phys. Rev.*, **94,** 1084.

ROCHESTER, C. D., and BUTLER, C. C., 1953, *Rep. Progr. Phys.* (London: Physical Society), **16,** 364.

ROSSI, B., 1955, *Proc. of the Fifth Annual Conference on High Energy Physics*, University of Rochester, in the press.

SCHWINGER, J., 1949, *Phys. Rev.*, **74,** 1439, **75,** 651, **76,** 790.

SHAPIRO, A. M., LEAVITT, E. P., and CHEN, F. I., 1954, *Phys. Rev.*, **95,** 624.

STEINBERGER, J., PANOFSKY, W. K. H., and STELLER, J., 1950, *Phys. Rev.*, **78,** 802.

TOMONAGA, S., 1946, *Progr. Theor. Phys., Osaka*, **1,** 27.

WEISSKOPF, V. F., and WIGNER, E. P., 1930, *Z. Phys.*, **63,** 54.

YUKAWA, H., 1935, *Proc. Phys. Math. Soc., Japan*, **17,** 48.